scottish institute for research in economics

# SIRE DISCUSSION PAPER

SIRE-DP-2010-92

## Growth Rate Estimation in the presence of Unit Roots

Monojit Chatterji
Homagni Choudhury

University of Dundee

www.sire.ac.uk

# Growth Rate Estimation in the presence of Unit Roots

**Monojit Chatterji**
(University of Dundee, UK)

**Homagni Choudhury***
(University of Dundee, UK)

**Abstract**

This study addresses the issue of the presence of a unit root on the growth rate estimation by the least-squares approach. We argue that when the log of a variable contains a unit root, i.e., it is not stationary then the growth rate estimate from the log-linear trend model is not a valid representation of the actual growth of the series. In fact, under such a situation, we show that the growth of the series is the cumulative impact of a stochastic process. As such the growth estimate from such a model is just a spurious representation of the actual growth of the series, which we refer to as a "*pseudo growth rate*". Hence such an estimate should be interpreted with caution. On the other hand, we highlight that the statistical representation of a series as containing a unit root is not easy to separate from an alternative description which represents the series as fundamentally deterministic (no unit root) but containing a structural break. In search of a way around this, our study presents a survey of both the theoretical and empirical literature on unit root tests that takes into account possible structural breaks. We show that when a series is trend-stationary with breaks, it is possible to use the log-linear trend model to obtain *well defined* estimates of growth rates for sub-periods which are valid representations of the actual growth of the series. Finally, to highlight the above issues, we carry out an empirical application whereby we estimate meaningful growth rates of real wages per worker for 51 industries from the organised manufacturing sector in India for the period 1973-2003, which are not only unbiased but also asymptotically efficient. We use these growth rate estimates to highlight the evolving inter-industry wage structure in India.

***Keywords***: Growth Rate, CAGR, AAGR, Unit Root, Trend Stationary, Structural Breaks, Real Wages, Inter-Industry Wage Structure

***JEL Classification***: C12, C13, C18, C22, J31

**First draft: August 2010**
**This version: October 2010**

---

*Corresponding author: h.choudhury@dundee.ac.uk.

## 1. Introduction

When it comes to estimating growth rate of a sufficiently long time series, the most commonly used method is the method of least-squares. The least-square growth rate is estimated by fitting an exponential trend to a time series variable, or alternately, by fitting a linear trend line to the logarithmic values of the variable in the relevant period. In other words, growth rates over long periods of time are estimated by regressing the logarithm of the variable under consideration on a deterministic trend. This technique is very commonly used in the social sciences including economics and financial research. For instance, international organizations like the World Bank and OECD generally use this technique to estimate growth rates of various indicator variables for which a long time series is available. Least-squares growth rates are used in the World Bank publications when measuring trend-wise growth in economic variables such as GDP, and GNP per capita (World Bank Statistical Manual[1]). Similarly least-squares growth rates are used by the OECD whenever there is a sufficiently long time series to permit a reliable calculation (OECD, 2005).

However, the implicit assumption under the above setting that is rarely explicitly mentioned is that the logarithm of the variable under consideration is trend-stationary. Alternatively, this assumption states that the variable of interest contains no unit roots. This reflects the fact that any underlying shocks are temporary in nature as opposed to the existence of a unit root which would imply that the shocks are permanent. And it is for this reason, Baffes & Vallee (2003) emphasise the trend-stationary versus unit root dichotomy when using a log-linear trend model to estimate growth rate of a time series variable.

The issue of unit roots or the temporary versus permanent nature of shocks has been considerably addressed in the econometrics literature over the last three decades, especially following the seminal work of Nelson and Plosser (1982) on the stochastic behaviour of economic time series, which found that most macroeconomic variables have a univariate time series structure with a unit root. This led to a series of empirical analyses of the stationary properties of numerous macroeconomic variables as well as

---

[1] For more on growth rate estimation methodology used by the world bank, see (permanent URL at world bank site) : http://go.worldbank.org/6ZTES0VQQ0

theoretical explorations of alternative approaches to test the unit root hypothesis. Immediate studies [see, for example, Stulz & Wasserfallen (1985), Wasserfallen (1986) Campbell & Mankiw (1987a, 1987b), Cochrane (1988), Corbae & Ouliaris (1988), Phillips & Perron (1988) and Perron (1988), among others] generally reaffirmed the findings of Nelson and Plosser (1982). The most important implication of the unit root revolution, sparked by this strand of literature, is that under this hypothesis, random shocks have permanent effects on the long run level of macroeconomic time series'; that is the fluctuations are not transitory. These findings were challenged by Perron (1989), who argued that, in the presence of a structural break, most macroeconomic variables are not characterised by a unit root and that fluctuations are indeed transitory. Persistence arises only from large and infrequent shocks, like the Great Crash of 1929 and the oil price shock of 1973, which are not a realisation of the underlying data generating mechanism; and that the variables return to their deterministic trends after small and frequent shocks. In this sense, Perron considers these shocks as exogenous. He explains that the assumption of an exogenous shock resulting in a structural break (level break or trend break or both) is used as a device to remove the influence of these shocks from the noise function. Following this revolutionary development, there has been a growing and evolving literature that has focused on the unit root hypothesis in the presence of structural change that has ranged from endogenously determining a single break point to endogenously determining multiple break points as well as theoretical exploration of alternative approaches to the unit root hypothesis in the presence of structural breaks, some of which we will highlight in this paper.

While the stochastic behaviour of time series under the unit root hypothesis with structural breaks has been extensively researched for various macroeconomic variables, the theoretical implications of a unit root on growth rate estimation, in the presence of structural breaks, has not been considered explicitly and is relatively unexplored area of research. The aim of this paper is to therefore address this issue by highlighting the effect of a unit root on the underlying assumptions and restrictions of a typical growth model in the presence of structural breaks. We argue that when the natural logarithm of the variable under study contains a unit root and hence is non-stationary, the OLS based log-linear trend method to estimate growth rate of a time-series variable will potentially give an estimate that is not a valid representation of the actual growth of the series. In fact, in the presence of a unit root, the growth rate of a series is not *well defined* in the sense that there is no deterministic component of the growth in the series and that the growth in the series is purely the cumulative impact of a stochastic process. On the other hand, the statistical representation of a series as containing a unit root is not easy to separate from an alternative description which represents the series as fundamentally deterministic (no unit root) but containing a structural break. We reserve a more detailed exposition of this to section two below. At the same time, we demonstrate these theoretical implications by empirically estimating the growth rates of real wage per worker for production (blue collar) workers from 51 industries of the organised manufacturing sector in India from 1973-74 to 2003-04. The rest of the paper is structured as follows. Section 2 brings out the issues with growth rate estimation and structural changes in a time series in the presence of a unit root. Section 3 presents the data and our results on estimation of growth rates of real wage per worker for the 51 industries are presented in Section 4. Finally, in Section 5, we summarise our findings and present some concluding remarks.

## 2. Growth Rate Estimation in the presence of a Unit Root: Theoretical Issues and Empirical Implications

In this section, we address the theoretical issues concerning growth rate estimation using the method of least squares in the presence of a unit root. We first present the theoretical framework behind the estimation of growth rate using the least squares principle in subsection 2.1, followed by a discussion of the limitations of this methodology when there is a unit root in the time series under consideration in subsection 2.2. In subsection 2.3 we present a chronological development of the literature on testing the unit root hypothesis which ranges from the traditional and very commonly used Augmented Dickey-Fuller (hereafter, ADF) test to the more recent tests that allow for endogenous structural breaks and are more powerful than the traditional ADF test. This is followed by a discussion of the empirical implications of rejecting the unit root hypothesis in the presence of structural break(s) on using the method of least squares to estimate growth rate in subsection 2.4.

## 2.1 Growth Rate Estimation by the Method of Least Squares: A Theoretical Framework

Typically, the growth rate, $g$, of a variable, $Y_t$, between two adjacent periods $t$ and $t\text{-}1$, is computed as,

$$g = \frac{Y_t - Y_{t-1}}{Y_{t-1}} \tag{1}$$

We can re-write (1) as,

$$Y_t = (1+g)Y_{t-1} \tag{2}$$

Note that, $g$, is the compounded (*annual*) growth rate in $Y_t$, measured between two adjacent periods (*years*) of time. Now, if our variable $Y_t$ is a long time series, then, we can assume two situations, when we are estimating growth rate of the series. In the first situation, we can assume that $Y_t$ grows at a rate $g$ in all periods and in the second situation, we can assume that growth rates of $Y_t$ vary over time. The theoretical models/implications for growth rate estimation are quite different under the two assumptions and we present them in order in subsections 2.1.1 and 2.1.2 respectively.

### 2.1.1 Growth model under constant growth rate assumption

When we assume that the growth rate, g, of the time series under study, $Y_t$, is constant over time, what we actually mean is that the growth rate between the two consecutive time periods will be the same for all periods, i.e.,

$$g = \frac{Y_t - Y_{t-1}}{Y_{t-1}} = \frac{Y_{t-1} - Y_{t-2}}{Y_{t-2}} = \ldots = \frac{Y_1 - Y_0}{Y_0} \tag{3}$$

Now, under the above assumption, if we subject $Y_t$ to stochastic shocks, $v_t$, in each period, then, it follows from (2), that,

$$Y_t = (1+g)Y_{t-1}v_t \tag{4}$$

Following Baffes and Vallee (2003), we assume that the error term, $v_t$, is log normally distributed with mean equal to 1. This is to reflect the proportionality nature of shocks. In other words, we assume a multiplicative log normal distribution of the shock term rather than an additive normal distribution. The reason behind this is quite simple- an additive normal distribution would imply similar magnitudes of shocks at all levels of $Y_t$, which then means that if $Y_t$ is growing over time, the relative importance of the stochastic part would fall to zero. Banerjee *et al.* (1993, pg 193) points out that if the error was an additive *i.i.d* $(0, \sigma^2)$ then it would imply that the absolute amount of growth varies around a fixed mean, and therefore that, as the series grows, the average amount of growth falls to zero as a proportion of the series itself. Moreover, $\sigma^2/var(Y_t)$ would tend to zero, forcing the series to become essentially deterministic in relative terms.

It follows, from (4), that $Y_{t-1} = (1+g)Y_{t-2}v_{t-1}$, which combined with (4) gives $Y_t = (1+g)^2 Y_{t-2}v_t v_{t-1}$. From further recursive substitution, we would then get,

$$Y_t = (1+g)^t Y_0 v_t v_{t-1} \ldots v_1 \qquad (5)$$

Taking natural logarithms on both sides of (5), we get,

$$\ln(Y_t) = \ln(Y_0) + t\ln(1+g) + \sum_t \ln(v_t) \qquad (6)$$

Now setting, $\alpha = ln(Y_0)$, $\beta = ln\,(1+g)$ and $\varepsilon_t = \sum_t \ln(v_t)$ in (6) gives,

$$\ln(Y_t) = \alpha + \beta t + \varepsilon_t \qquad (7)$$

The growth rate is calculated as $g = \exp(\beta) - 1$, where $\beta$ is typically estimated from (7) with OLS. For very small growth rates[2], $\beta$ is approximately equal to $g$ and hence the growth rate is often reported as the OLS estimate of $\beta$ from (7) rather than $g$. [3] The use of equation (7) to estimate growth rates is very common among empirical researchers.

### 2.1.2 Growth model under time-varying growth assumption

The constant growth assumption in section 2.1.1 is very restrictive and in fact, as Altinay (2004) shows, it does not represent reality. Therefore, if we assume that the growth rates vary over time, i.e. the growth rate between two consecutive time periods is not the same in all periods, we have,

$$g_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}}, \quad g_{t-1} = \frac{Y_{t-1} - Y_{t-2}}{Y_{t-2}}, \quad \ldots, g_1 = \frac{Y_1 - Y_0}{Y_0} \qquad (8)$$

where $g_t$ is the growth rate in time $t$. Altinay (2004) argues that if we assume that a growth rate in one period is independent of the previous growth rates, and that growth in

---

[2] If $g$ is small, then by taking the Taylor expansion of log(1+g) and omitting higher powers of g, we have $\ln(1+g) \approx g$. Hence, $\beta = \ln\,(1+g) \approx g$.

[3] For example, if $\beta = 0.01$, then g = 0.01005, or alternatively, when $\beta$ implies 1% growth, g implies 1.005% growth. While, on the other hand, if $\beta = 0.01$ then g = 0.1052, or alternatively, when $\beta$ implies 10% growth, g implies 10.52% growth.

one period is the basis of the growth in the next period, then by rearranging (8), we get, $Y_t = (1 + g_t)Y_{t-1}$ implying that $Y_t$ follows a geometric random walk and from further recursive substitution, we get,

$$Y_t = Y_0 \prod_{t=1}^{t} (1 + g_t) \tag{9}$$

In order to be able to use OLS estimation technique on (9), Altinay (2004) suggests that (stochastic) time varying growth rates ($g_t$) are to be modelled in a way such that it has a deterministic component and a stochastic disturbance term. Hence, following him, we assume that time-varying growth rate can be represented by a constant deterministic growth ($1+\overline{g}$) times a stochastic component, the error term ($\omega_t$), that has a log normal distribution. i.e.,

$$(1 + g_t) = (1 + \overline{g})\omega_t \tag{10}$$

Using (10) in (9), we get,

$$Y_t = Y_0 (1 + \overline{g})^t \prod_{t=1}^{t} \omega_t \tag{11}$$

Now, taking the natural logarithm on both sides, (11) can be transformed to a linear model, as follows,

$$\ln(Y_t) = \ln(Y_0) + t\ln(1 + \overline{g}) + \sum_t \ln(\omega_t) \tag{12}$$

Now setting $\alpha = ln\ (Y_0)$, $\beta = ln\ (1+\overline{g})$ and $\varepsilon_t = \sum_t \ln(\omega_t)$ in (12) gives,

$$\ln(Y_t) = \alpha + \beta t + \varepsilon_t \tag{13}$$

which is exactly equivalent to (7)- the constant growth equation. The growth rate is calculated as $g = \exp(\beta) - 1$, where $\beta$ is typically estimated from (13) with OLS. Like in (7), for very small growth rates, $\beta$ is approximately equal to $g$ and hence the growth rate is often reported as the OLS estimate of $\beta$ from (13) rather than $g$. Note that, in this case, even though the underlying assumption for the model is that growth varies with time, the estimate of the growth rate from (13) gives us only the *constant deterministic* component of time varying growth rate.

### 2.1.3 Commonly used econometric specification for estimation of growth rate: the log-linear trend model

The discussions in sections 2.1.1 and 2.1.2 show that whether we start with a framework that assumes constant growth or time-varying growth, we end up with a specification for the estimation of growth rate of a time-series variable by OLS which corresponds to fitting a trend line to the natural logarithm of the time series under consideration, i.e.,

$$\ln(Y_t) = \alpha + \beta t + \varepsilon_t \qquad (14)$$

where OLS estimate of $\beta$ is the required growth rate of the time series variable $Y_t$. Equation (14) is also called the log-linear trend method for growth rate estimation and the estimate of growth rate by this method is equivalent, in theory and principle, to the compounded (annual) growth rate (CAGR)[4].

The use of the log-linear trend method based on the least-square principles to estimate growth rates, as presented in (14), is very common among empirical researchers and has almost become a norm in economic research. However an important implicit assumption that researchers generally make, when using this model is that the error term, $\varepsilon_t$, in (14) is independently and identically distributed (*i.i.d*) random variable with a zero mean and constant variance $\sigma^2$ , i.e., $\varepsilon_t \sim i.i.d\ (0, \sigma^2)$. This assumption is needed to ensure that the model presented by (14) is a trend-stationary process of which residuals are stationary around a deterministic trend. But what happens when this assumption does not hold? This is an issue that has not attracted too much attention in the economics literature. In fact this assumption does not hold in both (7) and (13), by construction, something that we will illustrate and address in section 2.2.

## 2.2 Limitations of the log-linear trend model for growth rate estimation when the series under study may contain a Unit Root

[4] CAGR is the geometric mean growth rate on an annualised basis. If we assume continuous compounding (i.e. the compounding period is infinitesimally small) of a series, then we have $Y_t = Y_0 e^{\beta t}$ , where $Y_0$ is the initial value of the series at $t$ = 0 and $\beta$ is the CAGR (also called exponential growth rate). Taking natural logarithms on both sides, we get $\ln(Y_t) = \ln(Y_0)+\beta t$, which is equivalent to (7) and (13). $\beta$ can be estimated by OLS and represents the CAGR of the series.

In order to highlight the limitations of the log-linear trend method, presented in (14), let us consider a situation, where the error term, $\varepsilon_t$, in (14) follows a first order autoregressive process, i.e.,

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t \tag{15}$$

with $\rho$ being the highest autoregressive root , and $u_t$ is an I(0) white noise process. In the case of $\rho = 1$, which means $\varepsilon_t$ follows a unit root process, (14) becomes,

$$\ln(Y_t) = \alpha + \beta t + \sum_t u_t \tag{16}$$

This shows that *ln*($Y_t$) is I(1), or another representation of a random walk with a drift. Thus, when $\rho = 1$, the error term in (14) is I(1) and hence the natural log of the series under study ($Y_t$) is non-stationary. Under such a situation, $\alpha$ cannot be consistently estimated by any method, and the OLS estimator of $\beta$ is no longer asymptotically efficient [Altinay (2004)].

Before we proceed any further, it is important to highlight the significance of this issue of ln($Y_t$) containing a unit root i.e. it is non-stationary. To what extent the estimate of growth rate using (14) reflects the stochastic or deterministic component of the equation depends crucially on whether ln($Y_t$) contains a unit root or not. If ln($Y_t$) contains no unit root [i.e $|\rho| < 1$ in (15), implying that $\varepsilon_t$ in (14) is I(0) and hence ln($Y_t$) is trend-stationary], then the series is trend-stationary and hence the estimate of growth rate from (14) is a valid representation of the actual growth of the series. If, on the other hand, ln($Y_t$) contains a unit root implying that the series is non-stationary, then the average constant growth rate as hypothesised in (14) is not *well defined*. The reason behind this is that, if ln($Y_t$) is a non-stationary process, then the error term, $\varepsilon_t$, in (14) is not well behaved in the sense that it is not an I(0) white noise process and hence it is the stochastic component (i.e. $\varepsilon_t$) of (14) that drives $Y_t$ rather than its deterministic component (i.e. $t$). And it is for this reason, when $\rho = 1$ in (15), which means that ln($Y_t$) is non-stationary, the growth in the series is purely the cumulative impact of a stochastic process, as shown in (16). In fact the error terms in (7) and (13), which are the basis of (14), follow a pure random walk process, which is clearly evident, when we re-write them as,

Error term in (7):

$$\varepsilon_t = \sum_{t=1}^{t} \ln(\nu_t)$$
$$\varepsilon_t = \sum_{t=1}^{t-1} \ln(\nu_t) + \ln(\nu_t) \qquad (17)$$
$$\varepsilon_t = \varepsilon_{t-1} + u_t$$

Error term in (13):

$$\varepsilon_t = \sum_{t=1}^{t} \ln(\omega_t)$$
$$\varepsilon_t = \sum_{t=1}^{t-1} \ln(\omega_t) + \ln(\omega_t) \qquad (18)$$
$$\varepsilon_t = \varepsilon_{t-1} + u_t$$

Since we had earlier assumed $\upsilon_t$ and $\omega_t$ to have log normal distributions, their logarithmic transformation is *NID* $(0, \sigma^2)$ and hence we denote both $\ln(\upsilon_t)$ and $\ln(\omega_t)$ as $u_t$ - an I(0) white noise process. This shows that, by construction, the log-linear trend model in (7) and (13) [and hence in (14)] has obvious problem in its error term which is just the accumulation of past and current logarithmic disturbances implying that the series is an I(1) process as shown in (17) and (18).

Under such a situation of growth in the series being purely the cumulative impact of a stochastic process, the OLS estimate of β from (14) would be like some spurious representation of growth, which actually does not exist. In the rest of the paper, we'll refer the estimate of growth rate in such a situation as "*pseudo growth rate*", which should be interpreted with caution.

Now, theoretically, it can be argued that the typical way around the non-stationarity problem, as noted above, is to consider a first differenced variant of the log-linear trend method in (14), which gives,

$$\Delta \ln(Y_t) = \beta + u_t \qquad (19),$$

where Δ is the first difference operator, $\beta$ is a constant and $u_t$ is an I(0) white noise process. The OLS estimator of $\beta$ in (19) is now asymptotically efficient and is our required estimate of the growth rate of the series [Altinay (2004)]. However, we argue that this first differenced variant of the log-linear trend method (hereafter, log-difference method) is not free from problems and it is important to point out three key limitations of this model. First, this method gives an estimate of growth rate (if it exists), which is equivalent, in theory and principle, to the conventional average (constant) annual growth rate (AAGR)[5], rather than the CAGR. But since AAGR is simply the (linear) average of the period-to-period growth rates, it is more vulnerable to outliers than the CAGR and hence if the growth rate is particularly high (or low) between two consecutive periods for some reason, then this is going to pull up (or down) the estimate of the (constant) average growth rate for the entire period. Second, the conventional statistics (like adjusted-$R^2$, *t*-ratios etc.) associated with (14) and (19) would be very different from each other and hence any further statistical inferences based on (14) and (19) may potentially give contrasting conclusions. Third, and most importantly, if the natural log of the time series variable under study contains a unit root and hence is non-stationary at the very first place, then we have already argued earlier that growth in the series is cumulative impact of a stochastic process and under such a situation, it is not clear what does the point estimate from (19) even mean and hence it should be interpreted with caution. In fact it again gives us a "*pseudo growth rate*" even though, this time, the estimate is asymptotically efficient. Theoretically, when a series is trend-stationary and hence growth rate exists, its estimate from the log-linear method (CAGR), as presented in (14), should be identical to that from the log-difference method (AAGR), as presented in (19) [Altinay (2004)]. And it is probably for this reason that inferences based on (19) might give contrasting conclusions to that based on (14), as pointed out above.

Thus, to summarise, when $\rho = 1$ in (15), growth rate does not exist in the true sense of the term. The OLS estimate of $\beta$ from the log-linear trend model in (14) is asymptotically inefficient and gives us a spurious estimate of CAGR, which we call as a "*pseudo growth rate*". Further, the OLS estimate of $\beta$ from the log-difference model in (19) is asymptotically efficient, but this time it gives us a spurious estimate of AAGR, which is also some kind of a "*pseudo growth rate*". On the other hand, when $|\rho| < 1$ in (15), then $\varepsilon_t$ is I(0) in (14), thus implying that $\ln(Y_t)$ is a trend-stationary process. Under such a situation, when a time series variable is stationary around a deterministic trend and hence growth rate exists, its estimate from the log-linear method (which gives the CAGR), as presented in (14), should be identical to that from the log-difference method

[5] The AAGR, in principle, is the average of the period-to-period (annual) growth rate of a time-series variable.

(which gives the AAGR), as presented in (19) [Altinay (2004)]. And it is for this reason, Baffes & Vallee (2003) highlights that the trend-stationarity versus unit root dichotomy is an important issue to be considered when estimating growth rate of a series using the log-linear trend model. They therefore emphasise that the performance of the growth regressions should not only be assessed by examining the conventional statistics (like adjusted-$R^2$, $t$-ratios) but also by examining whether the log of the variable under consideration contains unit roots or not. While testing for unit roots, they focus mainly on the error term [in (14)], where they explore its stationarity properties on the basis of conventional ADF tests as well as Monte Carlo simulations. However, we argue that a more straightforward approach is to follow a two-step procedure. First, determine the stationarity properties of $\ln(Y_t)$, where $Y_t$ is the time series variable under study. Second, if $\ln(Y_t)$ is trend-stationary, then proceed with estimation of growth rates using the log-linear trend model presented in (14), or alternatively, if $\ln(Y_t)$ is non-stationary, then proceed with estimation of growth rates using the log-difference model presented in (19). However, we emphasise that when $\ln(Y_t)$ is non-stationary, then the estimate of growth rate of the series from (19) is asymptotically efficient but it is actually the "*pseudo growth rate*" and hence should be interpreted with caution.

This shows that the testing of the stationarity of a time series variable is a pre-condition to the estimation of growth rates using a log-linear trend model as described above. However, under a two step procedure as suggested above, the empirical concern is about the type of the unit root tests used in testing whether the natural log of the variable under study contains a unit root or not, especially in the light of evidence of low power of the conventional and commonly used Augmented Dickey-Fuller test. Hence in the following subsection, we present a historical evolution of the literature on unit root tests where the more recently suggested tests take into account many theoretical and empirical issues of which the most important is allowing the presence of structural break(s) and are therefore better in power and size than the traditional and well established procedures to test the unit root hypothesis.

### 2.3 Unit Root Testing Procedures: A Historical Review of Literature

The most widely used test for stationarity is the Augmented Dickey-Fuller (hereafter, ADF) test based on the statistical methodology put forward by Dickey & Fuller (1979, 1981). Most of the studies that we have earlier discussed in Section 1, including the seminal work of Nelson & Plosser (1982) and some of the work that followed it like Stulz & Wasserfallen (1985), Wasserfallen (1986), among others, that re-confirmed their findings that most macroeconomic variables are non-stationary (i.e. has a unit root), depended on a statistical methodology that was based on the ADF test, which relies on the estimation of the following augmented equation,

$$\Delta X_t = \hat{\alpha} + \hat{\beta} t + \hat{\rho} X_{t-1} + \sum_{j=1}^{k} \hat{\delta}_j \Delta X_{t-j} + \hat{e}_t \qquad (20)$$

where $\Delta$ is the first difference, $X_t$ is the time series being tested, $t$ is the time trend variable, $k$ is the number of lags added to the model[6] to ensure that the residuals, $\hat{\varepsilon}_t$ are white noise[7]. Non-rejection of the null hypothesis (i.e. $H_0$: $\rho = 0$) implies that the series $X_t$ is non-stationary, whereas rejection of the null in favour of the alternate (i.e $H_1$: $\rho < 0$) indicates that the series is trend-stationary.

However, Perron (1989) challenged these findings by arguing that, in the presence of a structural break, the standard ADF test is biased towards the non-rejection of the null hypothesis. In order to overcome this, he developed a procedure, where he allowed a single exogenous (known) break in both the null and the alternate hypothesis. This method tests the null hypothesis that a given series $\{X_t\}_1^T$ has a unit root with a non-zero drift and that a one time structural change occurs exogenously at a time $T_B$ (where $1<T_B<T$) versus the alternative hypothesis that the series is stationary about a deterministic time trend with an exogenous change in the trend function at time $T_B$. Perron considers three different models under the above setting.

In the first model, which Perron calls as the "crash model" (Model A in Perron's original 1989 paper), the null hypothesis that the series contains a unit root allows for a once only exogenous change in the level (or intercept) of the series, whereas, under the alternate hypothesis of trend-stationarity, the model permits for a one time change in the

[6] Generally, the Schwarz Bayesian Criterion (SBC) or Akaike Information Criterion (AIC) is used to choose the optimal lag length or $k$.

[7] This means that $\hat{e}_t$ has zero mean and constant variance that is uncorrelated with $\hat{e}_s$ for $t \neq s$.

intercept of the trend function occurring at the break-point. These hypotheses are parameterised as follows,

*Model (A): The "Crash" Model:*

$H_0$: $X_t = \alpha + dD(TB)_t + X_{t-1} + e_t$ ,

$H_1$: $X_t = \alpha_1 + \beta t + (\alpha_2 - \alpha_1)DU_t + e_t$ ,

where $D(TB) = 1$ if $t = T_B + 1$, 0 otherwise;

$DU_t = 1$ if $t > T_B$, 0 otherwise; and

$A(L)e_t = B(L)v_t$

with $v_t$ ~i.i.d (0, $\sigma^2$), and $A(L)$ and $B(L)$ are $p^{th}$ and $q^{th}$ order polynomials in the lag operator $L$ respectively. The null hypothesis is characterised by: (i) a unit root, and, (ii) a once only exogenous change in the level (intercept) denoted by a dummy variable which takes the value one just after the time of the break. On the other hand, the alternate hypothesis is characterised by: (i) trend-stationarity, and, (ii) an exogenous change in the level of the trend function occurring at the break-point, $T_B$, the magnitude of which is represented by $\alpha_2$-$\alpha_1$. In order to test for a unit root with this model, Perron used a modified ADF methodology, which involved the following augmented regression,

$$\textit{Model (A):}\ X_t = \hat{\alpha} + \hat{\theta}DU_t + \hat{\beta}t + \hat{d}D(TB)_t + \hat{\gamma}X_{t-1} + \sum_{j=1}^{k}\hat{\delta}_j \Delta X_{t-j} + \hat{e}_t \qquad (21)$$

where, $k$ is the number of lags added to the model[8] to ensure that the residuals, $\hat{e}_t$ are white noise.

In the second model, which Perron calls as the "changing growth" model (Model B in Perron's original 1989 paper), an exogenous change in the rate of growth (or slope) of the series is allowed. The null and alternate hypotheses, under this model, are parameterised as follows,

*Model (B): The "Changing Growth" Model*:

$H_0$: $X_t = \alpha_1 + X_{t-1} + (\alpha_2 - \alpha_1)DU_t + e_t$ ,

$H_1$: $X_t = \alpha + \beta_1 t + (\beta_2 - \beta_1)DT_t^* + e_t$ ,

where $DU_t = 1$ if $t > T_B$, 0 otherwise;

[8] The number of extra regressors, $k$, is determined by a test of significance of the estimated coefficients, $\hat{\delta}_j$ .

$DT_t* = t - T_B$ , if $t > T_B$, 0 otherwise; and,

$A(L)e_t = B(L)v_t$

with $v_t$ ~i.i.d (0, $\sigma^2$), and A($L$) and B($L$) are $p^{th}$ and $q^{th}$ order polynomials in the lag operator $L$ respectively. The null hypothesis is characterised by: (i) a unit root, and, (ii) a drift parameter $\alpha$ changing exogenously from $\alpha_1$ to $\alpha_2$ at the time of the break, $T_B$; whereas, the alternate hypothesis is characterised by: (i) trend-stationarity, and, (ii) an exogenous change in the slope of the trend function, as represented by $\beta_2$-$\beta_1$, without any sudden change in the intercept, occuring at $T_B$. The augmented equation, based on the ADF methodology, for this model is,

$$\textit{Model (B):}\ X_t = \hat{\alpha} + \hat{\beta}t + \hat{\mu}DT_t * + \hat{\gamma}X_{t-1} + \sum_{j=1}^{k}\hat{\delta}_j \Delta X_{t-j} + \hat{e}_t \qquad (22)$$

where, again, $k$ is the number of lags added to the model to ensure that the residuals, $\hat{e}_t$ are white noise.

Finally, the third model (Model C in Perron's original 1989 paper) that Perron considers combines changes in both level (intercept) and growth (slope) of the trend function of the series. Under this model, the null and alternate hypotheses are parameterised as follows,

*Model (C): Combination of both "Crash" and "Changing Growth" Model:*

$H_0$: $X_t = \alpha_1 + X_{t-1} + dD(TB)_t + (\alpha_2 - \alpha_1)DU_t + e_t$ ,

$H_1$: $X_t = \alpha_1 + \beta_1 t + (\alpha_2 - \alpha_1)DU_t + (\beta_2 - \beta_1)DT_t + e_t$ ,

where $D(TB) = 1$ if $t = T_B + 1$, 0 otherwise;

$DU_t = 1$ if $t > T_B$, 0 otherwise;

$DT_t = t$, if $t > T_B$, 0 otherwise; and,

$A(L)e_t = B(L)v_t$

with $v_t$ ~i.i.d (0, $\sigma^2$), and A($L$) and B($L$) are $p^{th}$ and $q^{th}$ order polynomials in the lag operator $L$ respectively. Under this model, the null hypothesis is characterised by: (i) a unit root, (ii) a once only exogenous change in the level (intercept), and, (iii) an exogenous change in the drift parameter from $\alpha_1$ to $\alpha_2$ occurring at the break-point, $T_B$. On the other hand, the alternate hypothesis is characterised by: (i) trend-stationarity, and, (ii) a change in both the level (intercept) and growth (slope) of the trend function occurring at the break-point, $T_B$, the magnitudes of which are represented by $\alpha_2$-$\alpha_1$ and $\beta_2$-

$\beta_1$ respectively. The augmented equation, based on the ADF methodology, for this model is,

$$\textit{Model (C):}\ X_t = \hat{\alpha} + \hat{\theta} DU_t + \hat{\beta} t + \hat{\mu} DT_t^* + \hat{d} D(TB)_t + \hat{\gamma} X_{t-1} + \sum_{j=1}^{k} \hat{\delta}_j \Delta X_{t-j} + \hat{e}_t \qquad (23)$$

where, again, $k$ is the number of lags added to the model to ensure that the residuals, $\hat{e}_t$ are white noise.

In order to formally test for the presence of a unit root ($H_0$: $\hat{\gamma}^i = 1$, $i$ = Models A, B or C), Perron used the following test statistics, which he computed from (21), (22) and (23),

$$t_{\hat{\gamma}^i}(\lambda), \quad \text{where } i = \text{models (A), (B) or (C)} \qquad (24)$$

which he showed to depend on the ratio of the pre-break sample size to the total sample size, $\lambda = T_B/T$. Perron used (24) to test the unit root hypotheses by comparing the above test statistic (which is the standard $t$-statistic to test $\hat{\gamma}^i = 1$) with its critical value from the asymptotic distribution of (24) for a fixed $\lambda$. He derived the asymptotic distributions for these test statistics under the preceding null hypotheses and tabulated their critical values for a selected grid of $\lambda$ values. If the value of the test statistic is smaller than that of the critical value, then the null hypothesis of a unit root can be summarily rejected. Based on this methodology, Perron showed that, for 11 out of the 14 time series analysed by Nelson & Plosser (1982) and the post war real GNP series analysed by Campbell & Mankiw (1987a, 1988b), the unit root hypothesis can be rejected.

Perron's (1989) methodology, of allowing for a structural break while testing for a unit root in a time series, is considered as a pioneering contribution in the literature of the unit root hypothesis. However, in spite of its pioneering nature, Perron's (1989) methodology has been the subject of severe criticism. Christiano (1992) criticised Perron's known (or exogenous) assumption of the break date by arguing that it involves an element of "data mining". He showed that, by using fixed date critical values, if a systematic search for a break in a series is carried out, when the series is actually a unit root process without break then it would entail a test that would lead to possible size distortions. Similarly, Zivot & Andrews (1992) argue that Perron's assumption of known breakpoints is based on prior observation of the data and hence problems associated with "pre-testing" are applicable to his procedure. Since then, the ensuing literature addressed this problem of the known break date by adopting a completely agnostic approach, where a complete and systematic search is done to endogenously determine the break date. Some of these include Perron & Vogelsang (1992), Banerjee *et al*. (1992), Amsler & Lee (1995), Perron (1997), Lumsdaine & Papell (1997), Clemente *et al*. (1998), Lee & Strazicich (2001, 2003), among others. However, in spite of the similarity in the assumption of the endogenous break(s), there are considerable methodological differences among some of these tests. While a complete review of these methodological differences is beyond the scope of this paper, we present a discussion on some of the more important tests that are increasingly getting popular among empirical researchers. For a chronological review of the evolution of the literature on the unit root hypothesis along with a review of the theoretical issues in some of these methodologies, see Perron (2005) and Glynn *et al*. (2007).

Banerjee *et al*. (1992) considered rolling and recursive tests, where the former uses a sample of fixed length (which is much smaller than the full sample) that moves sequentially from some starting date to the end of the sample, while the latter considers a fixed starting date for all tests and increases the sample used from some minimum value to the full sample. However, Glynn *et al.* (2007) point out that their methodology uses non-sequential tests to determine the breaks using only sub-samples of data. The fact that their tests are based on sub-samples implies that not all the information in the data is used suggesting that such tests may be low in power.

Zivot & Andrews (1992) adopted an alternative strategy that is more closely related to the methodology of Perron (1989). They argue against Perron's case, where he allows his test statistic in (24) to depend on $\lambda$ (= $T_B/T$), which is assumed to be exogenous. They question this exogeneity assumption and instead consider the structural break as an endogenous occurrence. As a result, unlike Perron, they have a single model, under the null hypothesis, given by,

$$H_0: X_t = \alpha + X_{t-1} + e_t \tag{25}$$

Their null considers that the series $X_t$ is integrated without any exogenous break and therefore allows them to treat the selection of the breakpoint ($\lambda$) for the dummy variables in Perron's augmented equations (21)-(23) as the outcome of an estimation procedure intended to fit $X_t$ to a certain trend stationary representation. Thus, under their alternative hypothesis, $X_t$ is a trend-stationary process with a one time break in the trend function (intercept break in Model A, slope break in Model B, or both in Model C) occurring at an unknown point in time. Their aim is to estimate the point of break that gives the most weight to the trend-stationary alternative. In their actual procedure, the breakpoint is chosen using the test statistic in (24) in a way that gives the least favourable result for the null hypothesis (25), i.e. $\lambda$ is so chosen as to minimise the one sided *t*-statistic used for testing the presence of a unit root ($H_0$: $\hat{\gamma}^i = 1$, $i$ = *models (A), (B) or (C)*). If $\hat{\lambda}^i_{\inf}$ stand for such a minimising value for model *i*, then the test statistic is given by,

$$t_{\hat{\gamma}^i}[\hat{\lambda}^i_{\inf}] = \inf_{\lambda \in \Psi} t_{\hat{\gamma}^i}(\lambda), \quad i = \text{mod}\,els\ \text{(A),(B),(C)} \tag{26}$$

where $\psi$ is a set containing some pre-specified range for the break fraction, λ, and hence a closed subset of (0, 1). It is to be noted here that there is now a single model under the null, and hence the dummy variable $D(TB)_t$ used in Perron's augmented equations (21)-(23) are no longer required. Zivot & Andrews, following Perron, uses a modified ADF framework, where the augmented equations used are:

Model (A): $$X_t = \hat{\alpha} + \hat{\theta} DU_t(\hat{\lambda}) + \hat{\beta} t + \hat{\gamma} X_{t-1} + \sum_{j=1}^{k} \hat{\delta}_j \Delta X_{t-j} + \hat{e}_t \tag{27}$$

Model (B): $$X_t = \hat{\alpha} + \hat{\beta} t + \hat{\mu} DT_t^*(\hat{\lambda}) + \hat{\gamma} X_{t-1} + \sum_{j=1}^{k} \hat{\delta}_j \Delta X_{t-j} + \hat{e}_t \tag{28}$$

Model (C): $$X_t = \hat{\alpha} + \hat{\theta} DU_t(\hat{\lambda}) + \hat{\beta} t + \hat{\mu} DT_t^* + \hat{\gamma} X_{t-1} + \sum_{j=1}^{k} \hat{\delta}_j \Delta X_{t-j} + \hat{e}_t \tag{29}$$

where $DU_t(\lambda) = 1$ if $t > T\lambda$, 0 otherwise; and $DT_t^*(\lambda) = t - T\lambda$, if $t > T\lambda$, 0 otherwise. "*Hats*" are used on the $\lambda$ parameters in (27)-(29) to emphasise that they correspond to estimated values of the break fraction. The asymptotic distribution theory of their statistics in (26) and their critical values are tabulated and presented in Zivot & Andrews (1992). It is important to point out here that their work involves substantial methodological difference from that of Perron (1989). The null hypothesis that they consider is that of a unit root process that has no break, while, under the alternate hypothesis, they consider a stationary process with break. Perron (2005) points out that the result of this in an asymmetric treatment of the specification of the trend function under the null and alternate hypotheses. Vogelsang & Perron (1998) shows that a rejection of the null hypothesis may be due to the presence of a unit root process with breaking trend. They further note that, in practice the distortions may be very small, but nevertheless it remains as a problematic feature of this approach.

Perron & Vogelsang (1992) and Perron (1997) further extend the analysis of Zivot & Andrews (1992) by proposing a class of tests that allows for two different versions of the transition effects of a structural break. The first is called the “additive outlier” (AO) model which specifies that the change to the new trend function occurs instantaneously; while the second is called the “innovational outlier” (IO) model which specifies that the change to the new trend function occurs gradually. The distinction between the two is important not only because they assume different transition paths but also because the statistical procedures to test for unit roots are different for the two models. These tests mainly rely on two approaches in endogenising the choice of the break point, both of which requires estimation of Dickey-Fuller type regressions at all possible break dates. In the first approach, the break date is chosen by minimising the *t*-statistic across all possible regressions. The second approach involves choosing a break date that maximise (or minimises) a statistic that test the significance of one or more trend break coefficients [Perron & Vogelsang (1998)]. Perron & Vogelsang (1992) applied the AO and IO models for non-trending data, while Perron (1997) extended their use to trending data. However, these tests are not free from criticism. Perron himself has questioned the power of these tests. Furthermore, some authors have raised the issue of the trade off between the power of the test and the amount of information incorporated with respect to the choice of the break date [Glynn *et al*. (2007)].

Apart from their power, the single most important and common criticism that the procedures proposed by Banerjee *et al.* (1992), Zivot & Andrews (1992), Perron & Vogelsang (1992), Perron (1997) is that these procedures can capture only the single most significant break in a series. Lumsdaine & Papell (1997) argue that if more than one break exists in a series, then allowing for one endogenous break is insufficient and results in a loss of information. There is a growing belief that the failure to allow for multiple breaks can cause the non-rejection of the unit root null and hence allowing for the possibility of multiple breaks that are endogenously determined from the data will provide further evidence against the unit root hypothesis [see, for example, Ben- David *et al.* (2003), Maddala & Kim (2003)]. As a result of this, several recent studies have proposed procedures that address the issue of allowing multiple break points in their tests for unit roots. An extension of Zivot & Andrews (1992) is provided by Lumsdaine & Papell (1997) who outline a procedure to test the null hypothesis that there is a unit root in the series against the alternate that the series is trend stationary in the presence of two structural breaks (level or slope or both). Similarly, Clemente *et al.* (1998) carry out an extension of Perron & Vogelsang (1992) by considering the case where the variable under consideration exhibits two changes in its mean (intercept or level). Furthermore, Ohara (1999) extends the approach popularised by Zivot & Andrews (1992) to the general case with $m$ endogenous breaks. He also shows that if a unit root test taking into account $m_1$ changes in slope is performed on a series, which actually has $m_0$ changes, with $m_0 > m_1$, then the least squares estimate of $\hat{\gamma}$ converges to 1 thus providing theoretical evidence to the argument that a non-rejection of the unit root hypothesis may be due to a wrong specification of the deterministic components included as regressors, a point highlighted as a rule by Campbell & Perron (1991) [Ohara (1999)].

The unit root tests discussed above that allow for the possibility of one or multiple structural breaks that are determined endogenously have also drawn severe criticism recently. It has been argued that the critical values used in these tests may potentially be biased as they have been derived under the null hypotheses that does not allow for any break(s). As a result of this, there may be serious size distortions in the presence of a unit root with break(s) [Nunes *et al*. (1997)] and may lead to some loss of power [Perron (2005), Glynn *et al.* (2007)]. Most importantly, Lee & Strazicich (2003) point out that if break(s) are not considered under the null then its rejection does not necessarily imply absence of a unit root, but would imply a rejection of a unit root without break(s). Hence empirical studies employing this kind of procedures might erroneously conclude that rejection of the null indicates evidence of trend-stationarity with break(s), instead of accepting that the series is in fact only difference stationary with break(s). However, it is important to point out here that Perron's (1989) exogenous break unit root test differ from the endogenous break(s) unit root tests, discussed above, in the sense that he allowed for a possibility of a break under both the null and the alternate hypothesis. Lee & Strazicich (2003) argue that in Perron's test, this treatment was necessary and appropriate because if they would not have allowed the break under the null then the unit root test statistic would have diverged as the size of the break increased under the null. Similarly, Lee & Strazicich (2001) provide evidence that, in endogenous break(s) unit root tests, the assumption of no break under the null causes the test statistic to diverge. This might result in a significant rejection of the unit root null when in fact the data generating process (DGP) is a unit root with break(s) [Lee & Strazicich (2003)].

In order to overcome the problems associated with endogenous break(s) unit root test, Lee & Strazicich (2001, 2003, hereafter LS1 and LS2 respectively) proposed a minimum Lagrange Multiplier (LM) test that not only endogenously determines one (in the LS1 test) and two (in the LS2 test) structural breaks but also circumvents the above noted biases and potential spurious rejections of the unit root null. In the rest of the paper, we will only refer to the LS2 test as it allows two breaks and hence is more general in nature (However, our technical discussion on LS2 is also applicable to LS1, except that LS1 allows for only one break). The test statistic that they use, which is based on Schmidt & Phillips' (1992, hereafter, SP) LM unit root test methodology, is invariant to breakpoint nuisance parameters.[9] They consider a DGP based on the unobserved components model, as follows:

$$X_t = \delta' Z_t + e_t , \qquad e_t = \gamma e_{t-1} + \varepsilon_t \tag{30}$$

where $Z_t$ is a vector of exogenous variables and $\varepsilon_t \sim i.i.d\ N(0, \sigma^2)$. If Z = [1, t]′, then the DGP is the same as the no break case of the SP LM unit root test. However, LS2 allow for two structural breaks, which means two level shifts in Perron's (1989) Model (A), two slope breaks in Model (B)[10] and two level and two slope breaks in Model (C). As such Model (A) can be described by $Z_t = [1, t, D_{1t}, D_{2t}]'$, where $D_{jt} = 1$ for $t \geq T_{Bj} + 1$, $j = 1, 2$ and 0 otherwise; and Model (C) can be described by $Z_t = [1, t, D_{1t}, D_{2t}, DT_{1t}, DT_{2t}]'$,

[9] Assuming no breaks under the null hypothesis may be necessary in the two endogenous breaks unit root test proposed by Lumsdaine and Papell (1997) to make their test statistic invariant to breakpoint nuisance parameters. But this assumption is not required in the LM test as the distribution is invariant to breakpoint nuisance parameter (see, Amsler & Lee, 1995)

[10] They omit an explicit discussion on Model (B) arguing that it is commonly held that most economic time series can be described adequately by models (A) and (C). As such we also omit the discussion of the LM unit root test on Model (B), but emphasise that the test also works for model (B) and the asymptotic distributions can be easily extended to this model.

where $DT_{jt} = t - T_{Bj}$ for $t \geq T_{Bj} + 1$, $j = 1,2$ and 0 otherwise. Their DGP allow breaks under both the null ($H_0$: $\gamma = 1$) and the alternate ($H_1$: $\gamma < 1$) hypotheses, which are parameterised as follows:

*Model (A):*

$H_0$: $X_t = \alpha_0 + d_1 B_{1t} + d_2 B_{2t} + X_{t-1} + \upsilon_{1t}$

$H_1$: $X_t = \alpha_1 + \beta t + d_1 D_{1t} + d_2 D_{2t} + \upsilon_{2t}$

and *Model (C):*

$H_0$: $X_t = \alpha_1 + d_1 B_{1t} + d_2 B_{2t} + \theta_1 D_{1t} + \theta_2 D_{2t} + X_{t-1} + e_{1t}$

$H_1$: $X_t = \alpha_1 + \beta t + d_1 D_{1t} + d_2 D_{2t} + \mu_1 DT_{1t} + \mu_2 DT_{2t} + e_{2t}$

where $\upsilon_{1t}$, $\upsilon_{2t}$, $e_{1t}$ and $e_{2t}$ are stationary error terms, $B_{jt} = 1$ for $t = T_{Bj} + 1$, $j = 1,2$ and 0 otherwise. They then estimate the two break LM unit root test statistic, on the basis of the LM (score) principle, from the following augmented equation:

$$\Delta X_t = \delta' \Delta Z_t + \Phi \tilde{S}_{t-1} + u_t \qquad (31)$$

where $\tilde{S}_t = X_t - \tilde{\Psi}_v - Z_t \tilde{\delta}$, $t = 2,..,T$; $\tilde{\delta}$ are coefficients on the regression of $\Delta X_t$ on $\Delta Z_t$; $\tilde{\Psi}_v$ is given[11] by $X_1 - Z_1 \tilde{\delta}$; $X_1$ and $Z_1$ denote the first observation of $X_t$ and $Z_t$, respectively. It is important to point out here that the augmented equation for the testing regression (31) involves $\Delta Z_t$ instead of $Z_t$ so that $\Delta Z_t$ becomes $[1, B_{1t}, B_{2t}]$ in case of Model (A) and $[1, B_{1t}, B_{2t}, D_{1t}, D_{2t}]$ in case of Model (C), where $B_{jt} = \Delta D_{jt}$ and $D_{jt} = \Delta DT_{jt}$, $j = 1, 2$. Thus, under the null, $B_{jt}$ and $D_{jt}$ correspond to a once only change in the intercept (i.e., a jump) and a permanent change in drift respectively and while, under the alternate they correspond to a change in the intercept and a change in the trend respectively. The unit root null hypothesis to be tested is then given by $H_0$: $\Phi = 0$ and the LM test statistics used to test this hypothesis are:

$$\tilde{\rho} = T\tilde{\Phi} \qquad (32)$$

$$\tilde{\tau} = t - statistic \text{ for testing the null } \Phi = 0 \qquad (33)$$

---

[11] This parameterisation is based on the Schmidt & Phillips' (1992) methodology, where they consider a DGP as follows: $y_t = \psi + \xi t + X_t$, $X_t = \beta X_{t-1} + \varepsilon_t$ to derive a test statistic based on the LM (score) principle to test the unit root hypothesis. Amsler & Lee (1995) extend their testing methodology to allow for one exogenous structural break in level, while Lee & Strazicich (2003) further extend it to allow for two endogenously determined structural breaks (in both level and slope).

Lee & Strazicich (2003) show that the asymptotic null distribution of the two break LM unit root test for Model (A) is invariant to the location ($\lambda_j = T_{Bj}/T$, $j = 1, 2$) and magnitude of structural breaks. However, they show, that this invariance property does not strictly hold for Model (C), as the asymptotic null distribution of the endogenous break LM test somewhat depends on the location of the break ($\lambda_j$) even though not on the magnitude of the breaks. To endogenously determine the location of the two breaks ($\lambda_j$), the minimum LM unit root test uses a grid search as follows:

$$LM_{\rho} = Inf_{\lambda}\tilde{\rho}(\lambda) \tag{34}$$

$$LM_{\tau} = Inf_{\lambda}\tilde{\tau}(\lambda) \tag{35}$$

the asymptotic distributions of which, along with the critical values are respectively presented and tabulated in Lee & Strazicich (2003). Note that the critical values for Model (C) depend somewhat on the location of the breaks, which are tabulated for a selected grid of $\lambda_j$ values in their paper.

The above discussion shows that the literature on the unit root hypothesis in the presence of structural change has received a significant amount of attention in the recent past. While there is no general consensus on the most appropriate methodology to perform unit root tests [Glynn *et al* (2007)], we conclude, from the above discussion, that the Lee & Strazicich (2003) LM test is the most powerful test and we prefer this test over all the other tests. However, in this paper, we carry out a number of unit root tests in Section 4 for comparison purposes. We summarise the null and alternate hypotheses of the tests that we adopt in this paper in Table 1.

### 2.4 Unit Root and Structural Break(s): Implications for Growth Rate Estimation

Having seen the evolution of the literature on unit roots and the advantages and disadvantages of various unit root tests in the above subsection, we can conclude that the popular and very commonly used ADF test is not a very powerful test. In the presence of structural break(s) in a time series variable, the test is biased towards the non-rejection of the null thus suggesting that the series is difference stationary. Using such a test would lead us to believe that most series contain a unit root and hence non-stationary, when in reality the series might simply be characterised by a structural break, which the test would fail to take into account. Under such circumstances, we would be forced to believe that growth of the series is the cumulative impact of a stochastic process, as shown in section 2.2 and as such growth rate (in the true sense of the term) does not exist. In other words, using the log-linear trend model in (14) to estimate the growth rate of the series would result in an estimate of growth rate which would not be a valid representation of the actual growth rate and hence would give us a "*pseudo growth rate*". Further, such an estimate would be asymptotically inefficient. This would lead us to instead use the log-difference model in (19) in order to ensure that the estimate we obtain is asymptotically efficient. However, under such circumstances, the problem of using (19) is that since growth of the series is caused by a stochastic process (as shown by the ADF test), the asymptotically efficient estimate that we obtain from the log-difference model in (19) is still some form of a "*pseudo growth rate*". Thus, as a result the estimate of growth rate from (19) which gives AAGR would be much different than the estimate of growth rate from (14) which gives CAGR, even though, theoretically, we would expect them to be similar when the original series (natural log of the variable) is trend-stationary. But apart from this, the conventional statistics (like adjusted-$R^2$, $t$-statistic etc.) associated with (14) and (19) would be very different from each other, possibly due to the inability of the ADF test to account for structural breaks and wrongly concluding that the series is non-stationary thus suggesting that it is required to take the difference of the series in order to obtain an asymptotically efficient estimate. And hence, under such a situation, any further statistical inferences based on (14) and (19) may give contrasting conclusions. This clearly shows the pitfalls of using the conventional ADF procedure to test for a unit root.

Given the pitfalls of the ADF test, we turn to the more powerful unit root tests that allow for the possibility of structural break(s) (either exogenous or endogenous) in their testing procedure. When using this strand of tests, we can allow for structural break(s) in the series, based on Perron's (1989) original Model (C)[12] and if we can reject the null hypothesis of a unit root[13], we can carry out the estimation of growth rate using (14). However, in this case, we will have two (or three) estimates of growth rate for a single

[12] As model (C) is a more general model that allows for both level break and slope break, we prefer this model over models (A) and (B), which respectively allow for a level break and slope break.

[13] Depending on the test we use, the null may or may not allow for break(s) while the alternate allows for break(s) in all the tests. In particular, Perron (1989) allows for an exogenous break under both the null and alternate, Zivot & Andrews (1992) and Lumsdaine & Papell (1997) allow one and two endogenously determined break(s) respectively, only under the alternate, while Lee & Strazicich (2003) allows two endogenously determined breaks under both the null and alternate. See section 2.3 for more details.

series: one for the pre-break (or the pre- first break) sample and one for the post-break (or the post- first break/ pre- second break) sample (and one for the post- second break sample). But both (or all three) estimates would be a valid representation of the true growth of the two (or three) sub-samples of the entire period. This would be equivalent to fitting two (or three) separate trend lines, by ordinary least squares, to the two (or three) sub-samples of the time series respectively. In order to illustrate this, let us consider a time series $X_t = \ln(Y_t)$, where $t = 1, 2, \ldots, T$ and we are interested in estimating the growth rate of the variable $Y_t$. Now if $X_t$ does not contain any unit root then it is stationary and we can use (14) to obtain an unbiased and asymptotically efficient estimate of the growth rate of the series for the whole period. But, however, if the series is stationary only with a break say at $t = T_B$, where $1<T_B<T$, then we cannot use (14) to obtain any meaningful estimate of the growth rate for the whole period. However, we can use (14) to fit two separate trend lines by OLS to the two sub-periods, $t = 1, 2, \ldots, T_B$ and $t = T_B+1, \ldots, T$, as follows,

$$\ln(Y_t) = \alpha_1 + \beta_1 t + \varepsilon_{1t}, \quad t = 1,2,\ldots,T_B \tag{36}$$

$$\ln(Y_t) = \alpha_2 + \beta_2 t + \varepsilon_{2t}, \quad t = T_B + 1,\ldots,T \tag{37}$$

where $\beta_1$ and $\beta_2$ are unbiased and asymptotically efficient estimates of growth rates for the pre-break and post-break periods respectively, which are valid representation of the true growth of the series for these two sub-sample periods. Alternately, we can use dummy variables for the two different sub-samples, and estimate the growth rates for the two sub-periods by fitting a single equation, as follows,

$$\ln(Y_t) = \alpha_1 D_1 + \alpha_2 D_2 + \beta_1 D_1 t + \beta_2 D_2 t + u_t \tag{38}$$

where $D_j$ is a dummy variable which takes the value 1 in the $j$-th sub-period and 0 otherwise. The estimates of $\beta_1$ and $\beta_2$ are the required growth rates for the two sub-periods which are exactly equal to the estimates of $\beta 1$ and $\beta_2$ from (36) and (37). This example can also be extended to a series that is trend-stationary with multiple breaks.

## 3. Data and Data Treatment

This study draws data from the EPW Research Foundation (India) who has collated data from the Annual Survey of Industries (ASI), published by the Central

Statistical Organisation, Ministry of Statistics and Programme Implementation, Government of India. ASI provides reasonably comprehensive and reliable industrial estimates at a disaggregated level for the organised manufacturing sector in India. It covers the entire factory sector except factories under the control of Defence Ministry, Oil storage depots and technical training institutes. 'Factories' are those which are registered as such under 2m(i) and 2m(ii) of the Factories Act, 1948 which respectively relates to units which employ 10 or more workers with the aid of power and units which employ 20 or more workers without the aid of power. ASI carries out complete enumeration of large factories on a census basis, and the remaining on a sample basis, where 'large units' are defined as factories employing 50 or more workers with aid of power or 100 or more workers without the aid of power. The EPW Research Foundation has collected the primary data from the ASI and has made available continuous annual data on industries from 1973-74 to 2003-04 after carrying out concordance of different series, wherever necessary.

The data from EPW Research Foundation, used in this study, is for 51 industries defined at three digit level of the National Industrial Classification, 1998 (NIC-98) for the period 1973-74 to 2003-04 (hereafter, 1973 to 2003). A description of the three digit industry codes is presented in Appendix Table 1. This data is particularly interesting because it covers a long and continuous period of time which coincides with India's episode of substantial economic reforms in 1991 that marks the change in the policy regime from highly restrictive import substitution industrialisation (ISI) to a regime that was characterised by radical reforms of trade, industrial and foreign exchange policies. We therefore emphasise that 1991, which marks the change in the policy regime, potentially represents a structural break in our data. But given the fact that there were some half-hearted reforms in the 1980s, before they fully took off in the 1990s, we might also potentially have two structural breaks- one in the 1980s and one in the 1990s.

The paper draws raw data on wages paid to production workers (blue-collar workers) reported in current rupees along with data on total number of workers by industry. Using the consumer price index (CPI) for industrial workers (with base 1982) to deflate wages to workers, we arrive at real wages to workers which we further divide by the total number of workers to arrive at real wages per worker (i.e. annual average real wages by industry). The data on CPI used to deflate the nominal figures is taken from the Handbook of Indian Statistics, Reserve Bank of India, 2001.

## 4. Growth Rate Estimation of Real Wages in the presence of Unit Root: An Empirical Application

In this section, we use the Indian data on real wages per worker for production workers from 51 manufacturing industries from 1973 to 2003 to estimate and compare their growth rates over the whole period. The objective of this exercise is to explore the inter-industry wage structure in India and at the same time to illustrate the issues related with the estimation of growth rates by OLS when the series are non-stationary or are trend-stationary with or without break(s). We first estimate the growth rate of real wage per worker for each industry by using (14) without considering whether each series is stationary or not. These results, which are estimates of CAGR of real wage per worker for the 51 industries, are presented in Table 2A. We also present a summary of the conventional statistics in Table 2B, which, at the first instance, show that the regression equations for estimating growth rate of real wage per worker for the 51 industries performed quite satisfactorily: of the total of 51 regressions, in 26 cases (i.e. 50.98% of all cases) the adjusted-$R^2$ exceeded 0.70. Furthermore, 94.12% of the 51 cases of growth rate estimates are significant at the 5% level confirming that the majority of models exhibited adequate performance. However, since we haven't tested for the stationarity of the natural logarithim of real wage per worker for each industry, we cannot be certain that the estimates of growth rates presented in Table 2A are asymptotically efficient and that they are a valid representation of the true growth rate. In fact a close look at the Durbin-Watson statistic (see Tables 2A and 2B) highlight that in 39 out of the 51 regressions (76.47% of all cases) there is evidence of positive serial correlation and in 52.94% of the total cases there is severe positive serial correlation[14] thus suggesting that there is high persistence in the data, which may be due to the potential presence of a unit root. Hence we conduct an ADF test on each series and find that the natural logarithm of real wages per worker is trend-stationary only for 21 out of 51 industries. This shows that 30 of the estimates of growth rates (58.82% of all cases) presented in Table 2A are not only

[14] As a rule of thumb, a DW statistic less than 1 is a cause of concern.

asymptotically inefficient but are also some kind of "*pseudo growth rates*". This result is consistent with the view that in the presence of unit roots, trends appear to be statistically significant much more often than stationarity tests would imply [Phillips (1986)], as shown in Table 2B. We therefore retain the estimates of growth rate of real wages per worker based on the log-linear trend model presented in (14) for the 21 industries, for which the series are trend-stationary and use the log-difference model presented in (19) to estimate the growth rates for the rest of the 30 industries, for which the series are non-stationary. Note that using (14) gives us estimates of CAGR, while (19) gives estimates of AAGR. However the AAGR estimates of real wages per worker for the 30 industries should be interpreted with caution as they still represent some kind of meaningless "*pseudo growth rate*", even though they are now asymptotically efficient. We present the ADF test results and the CAGR and AAGR estimates in Table 3. Theoretically, when a series is trend stationary, then the growth rate estimate from the log-linear trend model in (14) which gives CAGR should be identical to that from the log-difference model in (19) which gives AAGR, a point that we highlighted earlier. In Table 3, we can see that the growth rate estimates from both the methods yield very similar results for the series which are trend-stationary, while that for the series which are non-stationary, the growth rate estimates from the two methods differ significantly for majority of the cases.

But given the criticism of the power of the ADF test, due to its inability to account for any existing structural breaks, as emphasised in section 2.3, we test whether the ADF tests used above were biased because possible structural break(s) were ignored. We first consider Perron's (1989) test where he allows for one exogenous break (intercept break, slope break or both). This test has also been criticised for its assumption of an exogenous (or known) break, whereby it has been argued that such an assumption is based on prior observation of the data and hence problems associated with "pre-testing" are applicable to the test. However, we argue that such an assumption of an exogenous break is relevant to our data on real wages for the 51 Indian industries as India witnessed a substantial change in her policy regime in 1991, as pointed out in section 3. Prior to 1991 India was a staunch believer of inward looking import substitution industrialisation (ISI) policies, while in 1991 there was a radical change in outlook resulting in the adoption of a massive economic reforms program, of which trade liberalisation was an important component. We therefore emphasise that 1991, which marks the change in the policy regime, potentially represents a structural break in our data and hence we use Perron's (1989) methodology to test the unit root hypothesis allowing for an exogenous break in 1991.

But, we still recognise the criticism of the exogeneity assumption of the break in Perron's (1989) test that it involves an element of 'data mining'. Furthermore, there has been recent evidence that the reforms in India have actually been initialised much before 1991 and that the effect of these reforms were not felt immediately but with a lag [see, for example, Das (2001, 2003), Pangariya (2004, 2007), Choudhury (2007), Sen (2009)] and hence the break date might not coincide with the year in which the major reforms were initialised. Given this, we also consider Zivot & Andrews' (1992) and Lumsdaine & Papell's (1997) tests for unit roots, that allow for one and two endogenously determined breaks respectively. But given that these tests do not allow for break(s) under the null, they might result in spurious rejection of the unit root hypothesis and this therefore leads us to also consider Lee & Strazicich's (2003) test that allows for two breaks under both the null and the alternate hypothesis. We present the results[15] of all these unit root tests that allow for one or two exogenous or endogenous breaks along with the break dates in Table 4. Since our data is on real wages per worker, we follow Perron (1989), Zivot & Andrews (1992), Lumsdaine & Papell (1997) and Lee & Strazicich (2003) and assume Model C[16] (that allows for both level and slope breaks). In each test[17], we use lags of the dependent variable to correct for any serial correlation, and the optimal lag length ($k$) is determined by following the general-to-specific approach[18] used by Perron (1989) and

[15] For compactness, we just present the final results of the test along with the break date(s). For majority of the series, the $t$-statistics on estimated coefficients are significant at conventional levels of significance. Detailed output files (GAUSS files for Perron, Lumsdaine-Papell and Lee-Strazicich tests and STATA files for Zivot-Andrews test) are available from the authors upon request.

[16] We also tried Model A, but in majority of the cases, the $t$-statistics on estimated coefficients for Model C are significant at conventional levels of significance and hence we just present the results for Model C.

[17] We gratefully acknowledge Junsoo Lee for making the GAUSS codes used in Lee and Strazicich (2003) and the codes for the Perron (1989) and Lumsdaine-Papell (1997) tests freely available in his website: http://cba.ua.edu/~jlee/gauss . We had to slightly modify these codes for our purpose and had to fix a bug in the code for the Perron test. These codes are available from the authors upon request. For the Zivot-Andrews (1992) test, we gratefully acknowledge Christopher Baum for making the STATA routine freely available at http://fmwww.bc.edu/repec/bocode/z/zandrews.ado .

[18] We use the general-to-specific approach for lag selection in the Perron (1989), Lumsdaine-Papell (1997) and Lee-Strazicich (2003) tests, while for the Zivot-Andrews (1992) test, we use the Schwartz Bayesian Information Criteria (BIC).

suggested by Ng and Perron (1997, 2001). Again, we follow the above cited literature and start from a maximum of $k$ = 8 lagged terms. If the $t$-statistic on the maximum lagged term is not significant at the 10% asymptotic normal value of 1.654, we reduce the maximum lagged terms by one and repeat the test until the maximum lagged term is significant or no lags are found. For the endogenous break(s) test, once the optimal $k$ is determined at the breakpoint (or each combination of two break points), the break(s) are determined where the endogenous one (or two) break(s) test statistic is at minimum. In order to eliminate end-points, we choose a 'trimming region' to search for the possible break(s) over the time interval [0.1T, 0.9T].

Table 4 shows that, when we introduce an exogenous break in 1991, the real wages per worker for 23 industries out of 51 industries are trend-stationary. This is slightly better than the ADF test results where 21 out of 51 series were trend stationary. But if we compare the results of the ADF test (Table 3) with that of the Perron test (Table 4) that allows for a structural break (both level and slope) in 1991 more carefully, we find that the real wage per worker for 15 industries which were not stationary according to the ADF test become trend-stationary when we allow for a break in 1991 using Perron's (1989) methodology. However, the real wage per worker for 13 industries which were earlier trend-stationary according to the ADF test are now shown to be non-stationary when we introduce a break in 1991.While the remaining 23 industries out of the 51 industries does not exhibit any change in the conclusions[19] of the ADF test, even when we allow for the break in 1991- 15 of the series which were non-stationary according to the ADF test also remain non-stationary according to Perron's test, while 8 of the series which were trend-stationary according to the ADF test are still shown to be trend-stationary by Perron's test. We present these comparisons in Table 5.

The implications of these findings on growth rate estimation is that we can now use the log-linear trend model of (14) to estimate growth rate (CAGR) for the 15 series which were shown to be non-stationary by the ADF test earlier but are now trend-stationary when we account for the break at 1991. However, we will have to use a

[19] Of course, the $t$-statistics on the estimated coefficients are different between the ADF test and Perron's test and hence the statistical level of significance at which the null is rejected or not will be different for the two tests, even though the final verdict on whether the series is stationary or not still remains the same.

dummy variable to distinguish the two periods[20]- the pre-break and the post-break periods- which will thus give us two growth rates for the two periods. These estimates of growth rates will be well defined and would hence be a valid representation of the true growth rates of the series for the two sub-periods. Further, the two sub-period growth estimates will also be unbiased and asymptotically efficient. These results are presented in Table 6. Furthermore, the CAGR estimates that we presented in Table 3 for the 21 industries for which the natural log of the real wage per worker were found to be trend-stationary by the ADF test, are still valid.

Turning back to Table 4, we can see that when we use unit root tests that allow for break(s) that are determined endogenously, we find that the log of the real wages per worker becomes stationary for a greater number of industries. Zivot-Andrews test, which allows for one endogenous break, find that the log of the real wages per worker is trend-stationary for 35 out of the 51 industries. Of course, in this case, the break dates are different for different industries and ranges from as early as 1978 to as late as 1998. Interestingly, the real wages per worker are trend-stationary for only 2 industries with a break in 1991[21], while for 11 industries, with a break in 1989 and still another 11 industries, with a break in 1998 (see, Table 4). Furthermore, when we allow for two endogenous breaks with the Lumsdaine-Papell test, we find that this time the number of industries for which the log of real wages per worker are trend-stationary goes up to 39. Since the second break is statistically significant for majority of the series[22], we conclude that it is reasonable as well as justifiable to account for two breaks when conducting unit root tests. The first break ranges from 1983 to 1996, while the second one ranges from 1987 to 2000 for the 51 industries. However, for 46 out of 51 industries, the first break occurs before the economic reforms of 1991, while for 41 out of 51 industries, the second break occurs after the policy-regime change of 1991. The implications of these results on growth rate estimation are same as discussed earlier. We can use (14) to estimate growth

---

[20] Or alternately, we can fit two separate trend lines to the two sub-samples of the entire period as discussed in section 2.4.

[21] Interestingly, these two industries were non-stationary with a break in 1991 according to Perron's test (see Table 4). This may be because, Perron (1989) considers a break under the null, while Zivot and Andrews (1992) does not allow for a break under the null and hence the critical values of the test statistics are different which gives different outcomes. Additionally, the critical values of Perron's test are dependent on the break fraction.

[22] Detailed GAUSS output files are available from the authors, upon request.

rates of the sub-periods for the series which are trend-stationary with break(s). However, we do not present the estimates of growth rates of real wages per worker here, as the Zivot-Andrews and Lumsdaine-Papell tests have drawn the criticism that they result in spurious rejection of the unit root hypothesis, as noted in section 2.3 and we therefore turn to the Lee-Strazicich test, which is supposedly one of the most powerful unit root tests to date, that allows for two endogenously determined breaks under both their null and alternate hypothesis.

When we use the Lee-Strazicich test, we find that the natural log of the real wages per worker is trend-stationary for 49 out of 51 industries, with two endogenously determined breaks, which are statistically significant for majority of the cases[23], and range from 1983 to 2000, as presented in Table 4. The first break occurs in the pre-liberalisation period (prior to 1991) for 47 industries, while the second break occurs in the post-reforms period (post 1991) for 42 industries. Theoretically, we expected to find stronger rejections of the null using the Lumsdaine-Papell test than with the Lee-Starzicich LM test, as noted in section 2.3. However, our results are in line with those of Lee & Strazicich (2003) in the sense that, in case of real wages from Nelson & Plosser's (1982) data set, they were able to reject the null at a smaller significance level, using their LM test, than that by Lumsdaine & Papell (1997), while for most of the other series from the same data set, Lee & Strazicich could not reject the null and wherever they did, they rejected it at a higher significance level than the Lumsdaine-Papell test.

Given the results of the Lee-Strazicich test, we can use (14) to estimate three sub-period growth rates of real wage per worker for each of the 49 industries, for which the log of real wages per worker was found to be trend-stationary with two endogenously determined breaks. We present these estimates in Table 7A which are a valid representation of the actual growth of the series in each of the three sub-periods in the sense that the growth in the series in each sub-period is driven by the deterministic component and not by the stochastic component of the trend function. Further, the estimates of growth are not only unbiased but also asymptotically efficient. In order to asses the performance of the 49 regression equations for the estimation of three sub-period growth rates for each of the 49 industries, we also present a summary of the

---

[23] Detailed GAUSS output files are available from the authors, upon request.

conventional statistics in Table 7B. The summary of conventional statistics in Table 7B confirm the validity of the growth rates in Table 7A by highlighting that the regression equations for the 49 industries performed quite satisfactorily: in 38 out of 49 industries (77.55% of all cases) the adjusted-$R^2$ is greater than 0.70. Similarly, of the 147 growth rate estimates (3 sub-periods times 49 industries), 56.46% are statistically significant at 10% or smaller levels of significance. Moreover, this time, we see that in 27 out of the 49 growth regressions (55.10% of all cases), there is no serial correlation as suggested by the Durbin-Watson statistic, although there is evidence of positive serial correlation in a meagre 6.12% of the cases and negative serial correlation in a meagre 2.04% of the cases. This thus suggests that we can use these well defined and valid estimates of growth rates (and their associated statistics) for the 49 industries to draw further statistical inferences and carry out economic analysis. In the case of our dataset, we can use these estimates of growth rates of average real wages to study the inter-industry wage structure for the organised manufacturing sector in India.

The growth rate (CAGR) estimates of real wages per worker for production workers by industry presented in Table 7A highlight a clear picture of the evolving inter-industry wage structure of India in both the pre- reforms (prior to 1991) and post- reforms (after 1991) periods. However, since the break dates occur at different points in time for different industries, it is not possible to highlight a comparison of all the industries together. While a complete review of the inter-industry wage structure is beyond the scope of this paper, we nevertheless highlight three key results based on the growth rate estimates in Table 7A. First, the impact of liberalisation in India has not been uniform across all the industries. The fact that the break dates occur at different time points for different industries (see Table 7A) highlight the differential impact of the reforms on different industries. In each industry, this amounts to imparting two (deterministic) shocks to a wage growth process that is intrinsically stationary. Hence we find, for each industry, wage growth varied across three sub-periods. Second, in the case of the 49 industries out of 51, for which the natural log of real wage per worker is trend-stationary with two endogenously determined breaks, the first break occurs in the pre-liberalisation period (prior to 1991) for 47 industries, while the second break occurs in the post-liberalisation period (post 1991) for 42 industries (see Table 7A). This result lends support to the conjectures that the reforms in India have actually been initialised much before 1991 and that the impacts of the radical reforms episode of 1991 were not felt immediately, but with a lag. Third, the CAGR estimates of the real wage per worker, presented in Table 7A, show that the inter-industry wage structure in India has changed a lot in the period 1973-2003 and that it provides some evidence that the inter-industry wage differences have become more pronounced in the post-reforms period. Out of the 49 industries, there are 26 industries for which both the break dates are identical to at least one other industry, which makes it possible for us to compare the performance of these industries in terms of real wages per worker. For example, the CAGR of real wages per worker for industries 171 (Spinning, weaving and finishing of textiles), 201 (Saw milling and planing of wood), 252 (Manufacture of plastic products) and 321 (Manufacture of electronic valves, tubes and other electronic components) for the sub-period 1973-87 are 2.03%, 2.02%, 3.84% and 4.68% respectively, while for the sub-period 1988-2000, the CAGRs are -1.11%, -1.40%, -0.03% and 2.16% respectively. The first sub-period, which coincides with a policy regime in the Indian economy that was characterised by restrictive inward looking ISI policies, saw strong positive growth in real wages per worker for each of the four industries with an average of 3.1% (and standard deviation of 0.013) thus implying that real wages for the average worker in each industry was going up at a decent rate. While, on the other hand, the second sub-period, which coincides with a policy regime that was characterised by radical reforms of trade and industrial policies, witnessed not only a slump in the growth but even negative growth rates for some industries, with an average of -0.09% (and standard deviation of 0.016), thus implying that the inter-industry wage structure deteriorated in the post-reforms period. Similarly, a close look at the CAGR estimates of different sub-periods for the other industries, in Table 7A, clearly shows that the inter-industry wage differentials have increased in India over time and provide some evidence that it has become more pronounced in the post reforms period.

We intend to extend this analysis in our future research in order to study the entire inter-industry wage structure for the manufacturing sector in India by using these well defined growth rates of real wages per worker, presented in this study. More importantly, it would be interesting to explore the reasons of such variation in the growth rates of real wages per worker between the Indian industries.

## 5. Summary and Conclusions

This study addresses the possible pitfalls of using the least-squares approach to the estimation of growth rate of a time series variable, when the series in not stationary. The usual way to estimate growth rate for a long time series is to fit an exponential trend to the series. In other words, the natural logarithm of the variable is regressed on a constant and a linear time trend, where the OLS estimate of the coefficient on the trend is the required growth rate of the variable, which in theory and principle is an estimate of the compounded (constant) growth rate. If the observations of the series are annual, this method will give the compounded annual growth rate (CAGR). This study brings out the disadvantages of using such a log-linear trend model, when the log of the variable under study contains a unit root, i.e. it is non-stationary. The study highlights that, in the presence of a unit root, the log-linear model would give an asymptotically inefficient estimate of the growth rate. Also, the presence of a unit root would imply that the growth in the series is not deterministic in nature but it is purely the cumulative impact of a stochastic process. Hence the growth rate estimate obtained from the log-linear trend method would not be well defined in the sense that it would not be a valid representation of the actual growth of the series. At most, it would give us a "*pseudo growth rate*", which would be like a spurious representation of the growth of the series. Under such circumstances, any analysis or further statistical inferences based on such estimates of growth rates will, undoubtedly, give biased conclusions that would be questionable.

The study therefore emphasises that a unit root test of the natural logarithm of the variable under study is a pre-requisite before using a log-linear trend model to estimate the CAGR. If the unit root test suggests that the time series variable is actually difference-stationary rather than trend-stationary, then the first difference of the log-linear trend model (or a log-difference model) would in theory and principle give the average of the period-to-period growth rate (average annual growth rate, or AAGR, if the observations are annual) of the series. However, even if this estimate would now be asymptotically efficient, it would still be some form of a "*pseudo growth rate*" as the log-difference model is simply a variant of the log-linear trend model and as such if the log of the series is non-stationary the growth in the series is driven by the stochastic component of the trend function and hence growth rate of the series is not *well defined*.

The study then highlights that the popularly used ADF test to test for unit roots is low in power due to its inability to account for structural breaks in a variable. As such, reliance on such a test would often lead us to use the first differenced variant of the log-linear trend method in order to obtain an asymptotically efficient estimate of the growth rate of variable. However, when using this method, while we would be under the impression that we have obtained the AAGR of the series, we would in reality be obtaining an asymptotically efficient estimate of some form of a "*pseudo growth rate*", which actually does not exist. Hence such an estimate should be interpreted with some degree of caution.

This led us to present a historical review of the literature on testing the unit root hypothesis. While the conventionally used ADF test has drawn immense criticism for its inability to capture structural breaks and hence its inability to reject the null that the series contains a unit root, there has been a continuous evolution in the nature of the unit root tests since the late 1980s. We have presented an in-depth discussion on this which has ranged from Perron's (1989) test that allows for an exogenous break until the more recent tests of Lee & Strazicich (2003) that is based on the LM principle and allows for multiple breaks in the series, which are determined endogenously.

Finally we present an empirical application of the above issues by estimating the growth rate of real wages per worker for production (blue collar) workers from 51 manufacturing industries in India from 1973 to 2003. We find that when we use the conventional ADF test, the log of the real wages per worker is trend-stationary for only 21 industries. This implies that we can use the log-linear trend model to estimate well defined growth rates (CAGR) of real wages per worker for the 21 industries. However, when we use the more recent unit root tests, which allow for structural breaks, we find that a greater number of the series are trend-stationary with breaks. Particularly, when we use Perron's (1989) test, with an exogenous break in 1991- a year which saw radical economic reforms in India- we find that the log of real wages per worker is trend-stationary with a break in 1991 for 23 industries only. When we use the Zivot-Andrews

(1992) test, which allows for one endogenously determined break, we find that 35 series are trend-stationary with break dates which are different for different industries. When we use Lumsdaine-Papell (1997) test, which allows for two endogenously determined breaks, we find that 39 series are trend-stationary with different break dates, where in majority of the cases, both the breaks are significant. However, we rely on the more superior unit root test proposed by Lee & Strazicich (2003) based on the LM principle, which finds 49 out of 51 series to be trend-stationary, with two endogenously determined breaks for each series. In other words, this translates, for each industry, to imparting to (deterministic) shocks to a wage growth process that is intrinsically stationary. Hence, we find, for each industry, that wage growth varied across three sub-periods. The results from these tests imply that we can use the log-linear trend model to obtain *well defined* growth rates (CAGR) for the sub-periods that would be valid representation of the actual growth of the series. We find that there is quite a bit of variation in growth rates of real wages per worker for different industries which has clear implications on the changing inter-industry wage structure in India. We intend to use these meaningful estimates of growth rate of real wages per worker for the 49 industries that we have presented in this study, which are unbiased and asymptotically efficient, to study the changing inter-industry wage structure in India as a next step to this study. We also plan to explore why such a variation exists.

It might also be a good idea to consider the theoretical issues affecting a log-linear trend model for estimation of growth rates when there is serial correlation in the residuals of the growth regressions. In particular, in this paper, we showed that when $|\rho| < 1$ in (15), $\varepsilon_t$ in (14) is I(0) and hence the series is trend-stationary and the OLS estimate of growth rate from (14) is a valid representation of the actual growth of the series. However, as long as $\rho \neq 0$, the serial correlation problem remains to exist, although the series is trend-stationary [Altinay (2003)]. We also saw that in 8% of all the growth regressions in Table 7, there is evidence of the presence of serial correlation. Hence, the problems of estimation of growth rate using the log-linear trend model does not only depend on the unit root versus trend-stationary dichotomy, but also on the degree of serial correlation, an issue that we intend to explore in our future work.

| Table 1: Null and Alternate Hypotheses under the different Unit Root Tests | | |
|---|---|---|
| **Null**: There is a unit root | | |
| **Unit Root Test** | **Alternate** | **Assumes break(s) under the null** |
| ADF | The series is stationary | No |
| Perron (1989) | The series is trend-stationary with one exogenous break | Yes |
| Zivot and Andrews (1992) | The series is trend-stationary with one endogenous break | No |
| Lumsdaine and Paell (1997) | The series is trend-stationary with two endogenous breaks | No |
| Lee and Strazicich (2003) | The series is trend-stationary with two endogenous breaks | Yes |

| Table 2A: CAGR of Real Wages Per Worker | | | | | | | |
|---|---|---|---|---|---|---|---|
| $Y_t = \alpha + \beta t + \varepsilon_t,\ t = 1973,...,2003\ \mathit{for\ each}\ i, i = 1\ \mathit{to}\ 51$ | | | | | | | |
| $Y_t$ is the natural logarithm of real wage per worker for industry *I* at time *t*, $\beta$ estimate is required growth rate | | | | | | | |
| **Industry Code** | **CAGR ( $\hat{\beta}$ )** | **Adj-R$^2$** | **DW** | **Industry Code** | **CAGR ( $\hat{\beta}$ )** | **Adj-R$^2$** | **DW** |
| **151** | 0.0278*** | 0.83 | 0.69 | **269** | 0.0171*** | 0.83 | 0.87 |
| **152** | 0.0268*** | 0.87 | 0.57 | **271** | 0.0248*** | 0.82 | 1.07 |
| **153** | 0.0258*** | 0.83 | 1.40 | **272** | 0.0240*** | 0.91 | 0.86 |
| **154** | 0.0423*** | 0.79 | 0.54 | **281** | 0.0391*** | 0.90 | 1.34 |
| **155** | 0.0251*** | 0.85 | 0.80 | **289** | 0.0178*** | 0.71 | 0.49 |
| **160** | 0.0176*** | 0.74 | 0.95 | **291** | 0.0202*** | 0.75 | 0.64 |
| **171** | 0.0076*** | 0.34 | 0.36 | **292** | 0.0230*** | 0.88 | 0.50 |
| **172** | 0.0104*** | 0.39 | 0.98 | **293** | 0.0163*** | 0.41 | 1.16 |
| **173** | 0.0168*** | 0.58 | 1.32 | **300** | 0.0076*** | 0.19 | 1.30 |
| **181** | 0.0125*** | 0.72 | 1.37 | **311** | 0.0530*** | 0.94 | 1.21 |
| **182** | 0.0161*** | 0.32 | 1.30 | **312** | -0.0111 | 0.03 | 0.27 |
| **191** | 0.0106*** | 0.47 | 0.73 | **313** | 0.0113*** | 0.46 | 0.48 |
| **192** | -0.0044** | 0.16 | 1.48 | **314** | 0.0120*** | 0.35 | 0.57 |
| **201** | 0.0126*** | 0.53 | 0.88 | **319** | 0.0008 | -0.03 | 0.52 |
| **202** | 0.0154*** | 0.79 | 1.53 | **321** | 0.0388*** | 0.91 | 1.25 |
| **210** | 0.0171*** | 0.75 | 0.68 | **323** | 0.0018 | -0.01 | 1.35 |
| **221** | 0.0334*** | 0.95 | 1.50 | **331** | 0.0250*** | 0.70 | 0.42 |
| **222** | 0.0155*** | 0.48 | 0.60 | **332** | 0.0244*** | 0.73 | 1.43 |
| **231** | 0.0217*** | 0.88 | 1.79 | **333** | 0.0278*** | 0.90 | 2.05 |
| **232** | 0.0289*** | 0.68 | 0.98 | **341** | 0.0368*** | 0.96 | 1.43 |
| **241** | 0.0249*** | 0.90 | 0.89 | **342** | 0.0155*** | 0.44 | 1.56 |
| **242** | 0.0106*** | 0.34 | 0.29 | **351** | 0.0120*** | 0.37 | 0.96 |
| **251** | 0.0151*** | 0.60 | 0.78 | **352** | 0.0141*** | 0.48 | 0.49 |
| **252** | 0.0198*** | 0.67 | 1.09 | **359** | 0.0245*** | 0.84 | 1.66 |
| **261** | 0.0311*** | 0.89 | 1.38 | **361** | 0.0273*** | 0.35 | 1.10 |
| | | | | **369** | 0.0129*** | 0.60 | 1.22 |

**Notes:**
1. ***, **, * denotes that the estimate is statistically significant at 1%, 5% and 10% level respectively.
2. Adj-R$^2$ and DW stands for Adjusted-R$^2$ and Durbin-Watson statistic (*d*) respectively.

**Source:** Author's Calculations based on ASI Data

**Table 2B: Performance of growth equations: A Summary of the Conventional Statistics**

| Adjusted – $R^2$ | | $t$ – ratios | | Durbin Watson ($d$) | |
|---|---|---|---|---|---|
| ≥ 0.90 | 13.73% | ≤ 1% | 92.16% | ≤ 0.5 | 13.73% |
| ≥ 0.80 | 35.29% | ≤ 5% | 94.12% | ≤ 1 | 52.94% |
| ≥ 0.70 | 50.98% | ≤ 10% | 94.12% | ≤ 1.363 ($d_L$) | 76.47% |

**Notes:**

1. This table summarises the information on the Conventional Statistics from Table 2A and reports the percentage of cases that fulfils the criteria. For example, 13.73% of the total number of cases (51 growth equations for 51 industries in Table 2A) has an adjusted-R2 that is greater than or equal to 0.90.
2. *t*-ratios refer to all 51 growth estimates (whole period growth rates for 51 industries). For example, 94.12% of the 51 estimates of growth rates are statistically significant at 10% or lower levels of significance.
3. $d_L$ is the lower bound critical value for Durbin Watson statistic for a model with 1 explanatory variable and 31 observations at 5% significance level. In 76.47% of the total number of cases (51 growth equations for 51 industries in Table 2A), there is positive autocorrelation. For 52.94% cases there is severe positive autocorrelation suggesting that there is persistence in the data and that growth is a result of a cumulative stochastic process.

**Source:** Author's Calculations based on Table 2A

**Table 3: ADF Test and CAGR & AAGR of Real Wages Per Worker**

ADF Test:

$H_0$: There is a unit root i.e the series is non-stationary or I(1)

$H_1$: There is no unit root i.e the series is trend-stationary or I(0)

$CAGR: Y_t = \alpha + \beta t + \varepsilon_t ; AAGR: \Delta Y_t = \beta + \varepsilon_t$ ; where $Y_t$ is the natural log of real wage per worker at time $t$ = 1973-2003 for each industry $i$ = 1-51, $\beta$ estimate is the required growth rate

| Industry Code | ADF test | CAGR | AAGR | Industry Code | ADF test | CAGR | AAGR |
|---|---|---|---|---|---|---|---|
| **151** | NS | 0.0278 | 0.0179 | **269** | NS | 0.0171 | 0.0142 |
| **152** | NS | 0.0268 | 0.0203 | **271** | TS*** | 0.0248 | 0.0259 |
| **153** | TS** | 0.0258 | 0.0203 | **272** | NS | 0.0240 | 0.0253 |
| **154** | NS | 0.0423 | 0.0204 | **281** | TS** | 0.0391 | 0.0365 |
| **155** | NS | 0.0251 | 0.0188 | **289** | NS | 0.0178 | 0.0129 |
| **160** | NS | 0.0176 | 0.0109 | **291** | NS | 0.0202 | 0.0141 |
| **171** | NS | 0.0076 | 0.0037 | **292** | NS | 0.0230 | 0.0201 |
| **172** | NS | 0.0104 | 0.0059 | **293** | TS* | 0.0163 | 0.0125 |
| **173** | NS | 0.0168 | 0.0091 | **300** | TS** | 0.0076 | 0.0115 |
| **181** | TS*** | 0.0125 | 0.0082 | **311** | TS* | 0.0530 | 0.0521 |
| **182** | TS** | 0.0161 | 0.0243 | **312** | NS | -0.0111 | -0.0133 |
| **191** | NS | 0.0106 | 0.0065 | **313** | NS | 0.0113 | 0.008 |
| **192** | TS** | -0.0044 | -0.0049 | **314** | NS | 0.0120 | 0.0112 |
| **201** | NS | 0.0126 | 0.0092 | **319** | NS | 0.0008 | 0.0109 |
| **202** | TS*** | 0.0154 | 0.011 | **321** | TS*** | 0.0388 | 0.0287 |
| **210** | NS | 0.0171 | 0.0122 | **323** | TS** | 0.0018 | 0.0036 |
| **221** | TS** | 0.0334 | 0.0292 | **331** | NS | 0.0250 | 0.0161 |
| **222** | NS | 0.0155 | 0.0076 | **332** | TS** | 0.0244 | 0.0175 |
| **231** | TS*** | 0.0217 | 0.0169 | **333** | TS*** | 0.0278 | 0.0257 |
| **232** | NS | 0.0289 | 0.0211 | **341** | TS*** | 0.0368 | 0.0337 |
| **241** | NS | 0.0249 | 0.0241 | **342** | TS** | 0.0155 | 0.0122 |
| **242** | NS | 0.0106 | 0.0058 | **351** | NS | 0.0120 | 0.0094 |
| **251** | NS | 0.0151 | 0.0109 | **352** | NS | 0.0141 | 0.0112 |
| **252** | NS | 0.0198 | 0.0126 | **359** | NS | 0.0245 | 0.0233 |
| **261** | TS** | 0.0311 | 0.0217 | **361** | TS* | 0.0273 | 0.0278 |
| | | | | **369** | TS* | 0.0129 | 0.0129 |

Total Number of TS series: 21

**Notes:**

1. NS and TS stands for Non-stationary and Trend-stationary respectively.
2. In a trend stationary series, the star stands for the statistical level of significance at which the Null Hypothesis that the series contains a unit root is rejected; ***, **, * denotes significance levels of 1%, 5% and 10% respectively.
3. When a series is NS, then growth rate does not exist. The estimate we obtain is a "*pseudo growth* rate", which is not a valid representation of the actual growth rate- the CAGR estimate from the log-linear trend model is asymptotically inefficient and the AAGR estimate from the log-difference model is asymptotically efficient. We nevertheless report both for the sake of comparison and completeness. See section 2.2 for more details.
4. When a series is TS i.e I(0) then its first difference is also I(0). Hence we can also estimate the AAGR of the series. Theoretically, if the series is TS, then the growth rate estimate should be same from the log-linear method (CAGR) and log-difference method (AAGR). See section 2.4 for more details.

**Source:** Author's Calculations based on ASI Data

| Table 4: Unit Root Tests with Break(s) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Industry Code | Perron | Zivot-Andrews | | Lumsdaine-Papell | | | Lee-Strazicich | | | |
| | Result (Break at 1991) | Result | Break Date | Result | Break Dates | | Result | Break Dates | |
| **151** | NS | TS*** | 1989 | NS | 1983 | 1991 | TS** | 1985 | 1997 |
| **152** | TS*** | TS*** | 1995 | TS*** | 1984 | 1992 | TS*** | 1983 | 1989 |
| **153** | TS*** | TS** | 1981 | TS*** | 1984 | 1999 | TS*** | 1983 | 1990 |
| **154** | NS | NS | 1997 | TS*** | 1983 | 1994 | TS*** | 1987 | 1999 |
| **155** | TS*** | TS** | 1989 | TS*** | 1988 | 1999 | TS*** | 1985 | 1996 |
| **160** | NS | TS** | 1984 | TS*** | 1983 | 1994 | TS*** | 1983 | 1990 |
| **171** | NS | NS | 1989 | NS | 1987 | 2000 | TS*** | 1987 | 2000 |
| **172** | TS**^ | NS | 1990 | TS*** | 1984 | 1995 | TS*** | 1984 | 1992 |
| **173** | NS | NS | 1990 | TS** | 1985 | 1994 | TS*** | 1984 | 2000 |
| **181** | NS | TS*** | 1980 | TS*** | 1988 | 1999 | TS*** | 1983 | 1986 |
| **182** | NS | TS* | 1998 | NS | 1986 | 2000 | TS** | 1983 | 1996 |
| **191** | TS* | NS | 1990 | TS*** | 1985 | 1994 | TS*** | 1986 | 1995 |
| **192** | NS | TS*** | 1979 | TS*** | 1985 | 1988 | TS* | 1983 | 1994 |
| **201** | TS*** | TS*** | 1988 | TS*** | 1992 | 2000 | TS*** | 1987 | 2000 |
| **202** | TS* | TS*** | 1997 | TS*** | 1990 | 1999 | TS* | 1986 | 1998 |
| **210** | TS* | TS* | 1989 | NS | 1987 | 1991 | TS** | 1985 | 1997 |
| **221** | TS*** | TS*** | 1989 | TS*** | 1986 | 2000 | TS*** | 1985 | 1995 |
| **222** | TS*** | TS** | 1998 | TS*** | 1990 | 1997 | TS*** | 1986 | 1998 |
| **231** | NS | TS*** | 1994 | NS | 1984 | 2000 | TS*** | 1984 | 2000 |
| **232** | TS**^ | TS** | 1984 | TS* | 1985 | 1995 | TS*** | 1988 | 2000 |
| **241** | TS**^ | TS*** | 1978 | TS*** | 1992 | 1995 | TS*** | 1983 | 1994 |
| **242** | TS*** | NS | 1989 | TS*** | 1988 | 1995 | TS*** | 1993 | 1996 |
| **251** | NS | TS** | 1991 | TS*** | 1996 | 2000 | TS*** | 1991 | 2000 |
| **252** | TS* | TS*** | 1989 | NS | 1983 | 2000 | TS*** | 1987 | 2000 |
| **261** | TS*** | TS* | 1998 | TS*** | 1986 | 1997 | TS*** | 1985 | 1998 |
| **269** | NS | TS** | 1993 | TS* | 1988 | 1993 | TS*** | 1983 | 1997 |
| **271** | NS | TS*** | 1991 | NS | 1987 | 1994 | TS*** | 1986 | 1995 |
| **272** | NS | NS | 1997 | NS | 1983 | 1992 | TS*** | 1992 | 1999 |
| **281** | NS | NS | 1995 | TS*** | 1985 | 1990 | TS* | 1984 | 1997 |
| **289** | NS | NS | 1993 | TS*** | 1986 | 1991 | TS*** | 1983 | 1989 |
| **291** | NS | TS*** | 1998 | NS | 1988 | 1992 | TS*** | 1986 | 2000 |
| **292** | TS*** | TS*** | 1998 | TS*** | 1985 | 1988 | TS*** | 1989 | 2000 |
| **293** | TS* | TS*** | 1989 | TS**^ | 1986 | 1996 | TS*** | 1990 | 1997 |
| **300** | NS | NS | 1983 | NS | 1989 | 1995 | NS | 1990 | 2000 |
| **311** | NS | TS*** | 1998 | TS*** | 1983 | 1995 | NS | 1987 | 1996 |

Continued on next page

| Table 4: Continued from previous page | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Industry Code** | **Perron** | **Zivot-Andrews** | | **Lumsdaine-Papell** | | | **Lee-Strazicich** | | |
| | **Result (Break at 1991)** | **Result** | **Break Date** | **Result** | **Break Dates** | | **Result** | **Break Dates** | |
| **312** | NS | TS*** | 1998 | TS*** | 1987 | 1993 | TS*** | 1983 | 1999 |
| **313** | NS | NS | 1987 | TS*** | 1983 | 1988 | TS* | 1984 | 2000 |
| **314** | NS | TS** | 1997 | TS*** | 1984 | 1997 | TS*** | 1983 | 1986 |
| **319** | NS | TS*** | 1990 | TS*** | 1983 | 1997 | TS*** | 1983 | 1988 |
| **321** | NS | TS*** | 1992 | NS | 1991 | 1996 | TS*** | 1987 | 2000 |
| **323** | TS* | TS*** | 1989 | TS**^ | 1987 | 1997 | TS*** | 1986 | 1989 |
| **331** | TS** | NS | 1998 | TS**^ | 1984 | 2000 | TS*** | 1988 | 2000 |
| **332** | NS | TS*** | 1989 | TS*** | 1986 | 2000 | TS*** | 1986 | 1999 |
| **333** | NS | TS*** | 1998 | TS*** | 1985 | 1992 | TS*** | 1992 | 2000 |
| **341** | NS | TS*** | 1998 | TS**^ | 1983 | 1988 | TS*** | 1986 | 1997 |
| **342** | NS | TS*** | 1988 | NS | 1984 | 1995 | TS*** | 1984 | 1994 |
| **351** | NS | NS | 1990 | TS*** | 1985 | 1988 | TS*** | 1984 | 1987 |
| **352** | TS* | TS** | 1998 | TS*** | 1992 | 2000 | TS*** | 1984 | 1999 |
| **359** | TS*** | NS | 1996 | TS*** | 1983 | 1987 | TS*** | 1989 | 2000 |
| **361** | TS*** | NS | 1989 | TS*** | 1984 | 1992 | TS*** | 1983 | 1986 |
| **369** | TS*** | NS | 1990 | TS*** | 1984 | 1997 | TS** | 1983 | 1991 |
| **Total TS Series** | 23 | 35 | | 39 | | | 49 | | |

**Notes:**
1. NS and TS stands for Non-stationary and Trend-stationary respectively.
2. In a trend stationary series, the star stands for the statistical level of significance at which the Null Hypothesis that the series contains a unit root is rejected; ***, **^, **, * denotes significance levels of 1%, 2.5%, 5% and 10% respectively.
3. For all the tests, the model that allows break(s) in both level(s) and slope(s) is considered. This corresponds to Perron (1989) Model C and Lumsdaine and Papell (1997) Model CC.
4. For compactness, we just report the final result of the concerned unit root test with the break dates. But the detailed output files are available from the authors upon request.
5. Perron, Lumsdaine-Papell and Lee-Strazicich tests are conducted in GAUSS and uses critical values as reported in Perron (1989), Lumsdaine-Papell (1997) and Lee and Strazicich (2003); while Zivot-Andrews test is conducted in Stata and uses critical values as reported in Zivot and Andrews (1992).
6. The critical values for the Perron (1989) test depend on the location of the exogenous break, $\lambda = T_B/T$.
7. The critical values for the Lee and Strazicich (2003) test for Model (C) somewhat depend on the location of the break, $\lambda = (T_{B1}/T, T_{B2}/T)$

**Source:** Author's Calculations based on ASI Data

| Table 5: Comparison of ADF test and Perron's test (with exogenous break at 1991) results | | | | |
|---|---|---|---|---|
| | **Different conclusions** | | **Same conclusions** | |
| **Industry Codes that are:** | **NS (ADF) and TS (Perron)** | **TS (ADF) and NS (Perron)** | **NS (ADF) and NS (Perron)** | **TS (ADF) and TS (ADF)** |
| | 152 | 181 | 151 | 153 |
| | 155 | 182 | 154 | 202 |
| | 172 | 192 | 160 | 221 |
| | 191 | 231 | 171 | 261 |
| | 201 | 271 | 173 | 293 |
| | 210 | 281 | 251 | 323 |
| | 222 | 300 | 269 | 361 |
| | 232 | 311 | 272 | 369 |
| | 241 | 321 | 289 | |
| | 242 | 332 | 291 | |
| | 252 | 333 | 312 | |
| | 292 | 341 | 313 | |
| | 331 | 342 | 314 | |
| | 352 | | 319 | |
| | 359 | | 351 | |
| **Total No. of series** | **15** | **13** | **15** | **8** |

**Notes:**
1. NS and TS stand for non-stationary and trend-stationary.
2. The variable in question here is the natural log of real wages per worker

**Source:** Author's Calculations based on ASI Data

| **Table 6: Growth rate estimates of real wage per worker for the 15 industries presented in column 2, Table 5** | | | | |
|---|---|---|---|---|
| $Y_t = \alpha_1 D_1 + \alpha_2 D_2 + \beta_1 D_1 t + \beta_2 D_2 t + \varepsilon_t$, **where** $Y_t$ **is the natural log of real wage per worker at** $t$ **= 1973,..., 2003 for each industry** $i$ **= 1, 2, ..., 15;** $D_j$ **(**$j$ **= 1, 2) is a dummy variable which takes the value 1 in the** ***j-th*** **sub-period and 0 otherwise. OLS estimate of** $\beta_j$ **(**$j$**=1, 2) is the required growth rate for the** ***j-th*** **sub-period** | | | | |
| **Industry Codes** | **1973-1991** | **1992-2003** | **Adj-$R^2$** | **DW** |
| **152** | 0.0386*** | 0.0024 | 0.95 | 1.65 |
| **155** | 0.0339*** | 0.0028 | 0.9 | 1.42 |
| **172** | 0.02440*** | -0.0159** | 0.72 | 2.30 |
| **191** | 0.02061*** | -0.0160** | 0.73 | 1.57 |
| **201** | 0.0254*** | -0.0085 | 0.75 | 1.79 |
| **210** | 0.0265*** | -0.0039 | 0.86 | 1.33 |
| **222** | 0.0276*** | -0.0206** | 0.69 | 1.14 |
| **232** | 0.0241*** | 0.0335** | 0.65 | 1.00 |
| **241** | 0.0343*** | 0.0160*** | 0.93 | 1.63 |
| **242** | 0.0309*** | -0.0227*** | 0.88 | 1.80 |
| **252** | 0.0357*** | -0.0038 | 0.83 | 2.31 |
| **292** | 0.0325*** | 0.0015 | 0.95 | 1.57 |
| **331** | 0.0395*** | -0.0259*** | 0.92 | 1.81 |
| **352** | 0.0317*** | -0.0171** | 0.82 | 1.46 |
| **359** | 0.0361*** | 0.0151** | 0.89 | 2.70 |

**Notes:**
1. ***, **, * denotes that the estimate is statistically significant at 1%, 5% and 10% level respectively.
2. Adj-$R^2$ and DW stands for Adjusted-$R^2$ and Durbin-Watson statistic ($d$) respectively.

**Source:** Author's Calculations based on ASI Data

| Table 7A: Growth rate estimates of real wage per worker for all industries | | | | | | | |
|---|---|---|---|---|---|---|---|
| $Y_t = \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \beta_1 D_1 t + \beta_2 D_2 t + \beta_3 D_3 t + \varepsilon_t$, **where $Y_t$ is the natural log of real wage per worker at $t$ = 1973,..., 2003 for each industry $i$ = 1, 2, ..., 15; $D_j$ ($j$ = 1, 2) is a dummy variable which takes the value 1 in the *j-th* sub-period and 0 otherwise. OLS estimate of $\beta_j$ ($j$=1, 2, 3) is the required growth rate for the *j-th* sub-period** | | | | | | | |
| **Industry Codes** | **Break Points** | | ($\hat{\beta}_1$) | ($\hat{\beta}_2$) | ($\hat{\beta}_3$) | **Adj-$R^2$** | **DW** |
| | **$T_{B1}$** | **$T_{B2}$** | **1973 - $T_{B1}$** | **($T_{B1}$+1)- $T_{B2}$** | **($T_{B2}$+1)-2003** | | |
| **151** | 1985 | 1997 | 0.0432*** | 0.0170** | 0.0210 | 0.93 | 1.83 |
| **152** | 1983 | 1989 | 0.0444*** | 0.0370*** | 0.0032 | 0.96 | 1.74 |
| **153** | 1983 | 1990 | 0.0390*** | 0.0401** | 0.0117 | 0.85 | 1.87 |
| **154** | 1987 | 1999 | 0.0593*** | 0.0092 | -0.0258 | 0.88 | 1.17 |
| **155** | 1985 | 1996 | 0.0319*** | 0.0346*** | 0.0187 | 0.91 | 1.58 |
| **160** | 1983 | 1990 | 0.0116* | 0.0185 | 0.0026 | 0.89 | 1.89 |
| **171** | 1987 | 2000 | 0.0203*** | -0.0111*** | -0.0192 | 0.81 | 1.38 |
| **172** | 1984 | 1992 | 0.0295*** | 0.0283** | -0.0145* | 0.72 | 2.30 |
| **173** | 1984 | 2000 | 0.0312*** | 0.0049 | -0.0070 | 0.74 | 2.41 |
| **181** | 1983 | 1986 | 0.0056 | 0.0445 | 0.0120*** | 0.78 | 1.65 |
| **182** | 1983 | 1996 | 0.0552*** | 0.0158 | 0.0673* | 0.43 | 2.07 |
| **191** | 1986 | 1995 | 0.0207*** | 0.0182* | -0.0146 | 0.70 | 1.53 |
| **192** | 1983 | 1994 | 0.0095 | -0.0183** | -0.0139 | 0.39 | 2.17 |
| **201** | 1987 | 2000 | 0.0202*** | -0.0140*** | 0.0191 | 0.81 | 2.02 |
| **202** | 1986 | 1998 | 0.0234*** | -0.0008 | -0.0119 | 0.85 | 2.01 |
| **210** | 1985 | 1997 | 0.0240*** | 0.0178*** | 0.0071 | 0.85 | 1.27 |
| **221** | 1985 | 1995 | 0.0414*** | 0.0316*** | 0.0177* | 0.96 | 2.23 |
| **222** | 1986 | 1998 | 0.0260*** | 0.0095 | 0.0250 | 0.73 | 1.37 |
| **231** | 1984 | 2000 | 0.0220*** | 0.0166*** | -0.0517 | 0.88 | 2.28 |
| **232** | 1988 | 2000 | 0.0164 | 0.0336** | -0.0236 | 0.65 | 1.08 |
| **241** | 1983 | 1994 | 0.0430*** | 0.0180*** | 0.0182** | 0.95 | 1.94 |
| **242** | 1993 | 1996 | 0.0279*** | -0.0257 | -0.0166 | 0.85 | 1.60 |
| **251** | 1991 | 2000 | 0.0297*** | -0.0157 | 0.0146 | 0.81 | 1.89 |
| **252** | 1987 | 2000 | 0.0384*** | -0.0003 | -0.0081 | 0.82 | 2.32 |
| **261** | 1985 | 1998 | 0.0280*** | 0.0259*** | 0.0276 | 0.90 | 1.88 |
| **269** | 1983 | 1997 | 0.0252*** | 0.0246*** | -0.0069 | 0.91 | 1.86 |
| **271** | 1986 | 1995 | 0.0227*** | 0.0269* | 0.0275 | 0.81 | 1.23 |
| **272** | 1992 | 1999 | 0.0315*** | -0.0031 | 0.0489** | 0.95 | 1.78 |
| **281** | 1984 | 1997 | 0.0359*** | 0.0419*** | 0.0439* | 0.91 | 1.57 |
| **289** | 1983 | 1989 | 0.0367*** | 0.0326** | -0.0073* | 0.92 | 1.90 |
| **291** | 1986 | 2000 | 0.0365** | 0.0082* | 0.0230 | 0.89 | 1.63 |
| **292** | 1989 | 2000 | 0.0324*** | 0.0062 | 0.0077 | 0.96 | 1.59 |
| **293** | 1990 | 1997 | 0.0398*** | 0.0040 | 0.0169 | 0.73 | 2.26 |
| **312** | 1983 | 1999 | 0.0490* | -0.0330** | -0.0177 | 0.61 | 1.12 |
| **313** | 1984 | 2000 | 0.0347*** | -0.0036 | -0.0435 | 0.88 | 2.34 |
| **314** | 1983 | 1986 | 0.0346*** | 0.0497 | -0.0103* | 0.65 | 1.28 |
| **319** | 1983 | 1988 | 0.0564** | 0.0252 | 0.0440*** | 0.36 | 1.46 |
| **321** | 1987 | 2000 | 0.0468*** | 0.0217** | -0.0082 | 0.91 | 1.63 |
| **323** | 1986 | 1989 | 0.0185*** | -0.1408** | 0.0120* | 0.28 | 2.00 |
| **331** | 1988 | 2000 | 0.0340*** | -0.0084 | -0.0482 | 0.90 | 1.48 |
| **332** | 1986 | 1999 | 0.0179* | 0.0282*** | -0.0410 | 0.74 | 1.68 |

Continued on next page

| Table 7A: Continued from previous page | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Industry Codes** | **Break Points** | | ($\hat{\beta}_1$) | ($\hat{\beta}_2$) | ($\hat{\beta}_3$) | **Adj-$R^2$** | **DW** |
| | **$T_{B1}$** | **$T_{B2}$** | **1973 - $T_{B1}$** | **($T_{B1}$+1)- $T_{B2}$** | **($T_{B2}$+1)-2003** | | |
| **333** | 1992 | 2000 | 0.0285*** | 0.0182 | -0.0265 | 0.90 | 2.16 |
| **341** | 1986 | 1997 | 0.0368*** | 0.0208*** | 0.0345** | 0.98 | 2.10 |
| **342** | 1984 | 1994 | 0.0195 | 0.0358** | -0.0221 | 0.48 | 1.97 |
| **351** | 1984 | 1987 | 0.0454*** | -0.0031 | 0.0101 | 0.55 | 1.64 |
| **352** | 1984 | 1999 | 0.0313*** | 0.0033 | -0.0285 | 0.81 | 1.59 |
| **359** | 1989 | 2000 | 0.0351*** | -0.0009 | 0.0367 | 0.90 | 2.96 |
| **361** | 1983 | 1986 | 0.0663** | 0.1550 | 0.0098 | 0.44 | 1.47 |
| **369** | 1983 | 1991 | 0.0270*** | 0.0009 | 0.0071 | 0.75 | 2.28 |

**Notes:**
1. ***, **, * denotes that the estimate is statistically significant at 1%, 5% and 10% level respectively.
2. This table present the growth rate estimates of real wage per worker for only those industries for which the natural log of real wage per worker is trend-stationary with two endogenously determined breaks on the basis of the Lee and Strazicich (2003) test (see Table 4)
3. Adj-$R^2$ and DW stands for Adjusted-$R^2$ and Durbin-Watson statistic (*d*) respectively.

**Source:** Author's Calculations based on ASI Data

| **Table 7B: Performance of growth equations: A Summary of the Conventional Statistics** | | | | | |
|---|---|---|---|---|---|
| **Adjusted – $R^2$** | | ***t* – ratios** | | **Durbin Watson (*d*)** | |
| $\geq 0.90$ | 28.57% | $\leq 1\%$ | 37.42% | $\leq 1.229$ ($d_L$) | 6.12% |
| $\geq 0.80$ | 63.27% | $\leq 5\%$ | 47.62% | $\geq 2.771$ ($4 - d_L$) | 02.04% |
| $\geq 0.70$ | 77.55% | $\leq 10\%$ | 56.46% | $1.650$ ($d_U$)$\leq d \leq 2.35$ ($4 - d_U$) | 55.10% |

**Notes:**

1. This table summarises the information on the Conventional Statistics from Table 7A and reports the percentage of cases that fulfils the criteria. For example, 28.57% of the total number of cases (49 growth equations for 49 industries in Table 7A) has an adjusted-$R^2$ that is greater than or equal to 0.90.
2. *t*-ratios refer to all 147 growth estimates (3 periods times 49 industries). For example, 56.46% of the 147 estimates of growth rates are statistically significant at 10% or higher levels of significance.
3. $d_L$ and $d_U$ are the lower and upper bound critical values respectively for Durbin Watson statistic for a model with 3 explanatory variables and 31 observations at 5% significance level. In 55.10% of the total number of cases (49 growth equations for 49 industries in Table 7A), we cannot reject the null hypothesis that there is no serial correlation. For 12.24% cases there is positive autocorrelation and for 2.04% of the cases, there is negative autocorrelation.

**Source:** Author's Calculations based on Table 7A

## References


Altinay, G. (2003), Estimating growth rate in the presence of serially correlated errors, Applied Economic Letters, Volume 10, Issue 15, December 2003, pages 967-970.

Altinay, G. (2004), On measuring average growth, Applied Economics, Volume 36, Issue 6, April 2004, pages 637 – 644.

Amsler, C. and J. Lee (1995), An LM Test for a Unit Root in the Presence of a Structural Change, Econometric Theory, Volume 11, Issue 2, June 1995, pages 359 – 368.

Annual Survey of Industries (ASI) 1973-74 to 2003-04: Volume II, EPW Research Foundation, Sameeksha Trust (Mumbai, India), January 2007.

Baffes, J. and J.L. Vallee (2003), Unit roots versus trend stationarity in growth rate estimation, Applied Economics Letters, Volume 10, Issue 1, January 2003 , pages 9 – 14.

Banerjee, A., R. L. Lumsdaine and J. H. Stock (1992), Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence, Journal of Business and Economic Statistics, American Statistical Association, Volume 10, Issue 3, July 1992, pages 271 – 287.

Banerjee, A., J. Dolado, J. W. Galbraith and D. Hendry (1993), Co-integration, Error Correction, and the Econometric Analysis of Non-Stationary Data, Oxford University Press, July 1993.

Ben-David, D., R. L. Lumsdaine and D. H. Papell (2003), Unit roots, postwar slowdowns and long-run growth: Evidence from two structural breaks, Empirical Economics, Springer, Volume 28, Issue 2, April 2003, pages 303 – 319.

Campbell, J. Y and N. G. Mankiw (1987a), Permanent and Transitory Components in Macroeconomic Fluctuations, The American Economic Review*,* Volume 77, Issue 2, Papers and Proceedings of the Ninety-Ninth Annual Meeting of the American Economic Association, May, 1987, pages 111 – 117.

Campbell, J. Y and N. G. Mankiw (1987b), Are output fluctuations transitory?, The Quarterly Journal of Economics, Volume 104, Issue 4, November 1987, pages 857 – 880.

Campbell, J. Y. and P. Perron (1991), Pitfalls and Opportunities: What Macroeconomists

Should Know About Unit Roots, NBER Chapters, in: NBER Macroeconomics Annual 1991, Volume 6, pages 141-220 National Bureau of Economic Research.

Choudhury (2007), The Impact of Trade Liberalization on Aspects of Economic Performance in the Organized Manufacturing Sector in India, Unpublished MSc Dissertation, Department of Economics, Lancaster University, UK, Sep 2007.

Christiano, L. J. (1992), Searching for a Break in GNP, Journal of Business and Economic Statistics, American Statistical Association, Volume 10, Issue 3, July 1992, pages 237 – 250.

Clemente, J., A. Montanes, and M. Reyes (1998), Testing for a unit root in variables with a double change in the mean, Economics Letters, Elsevier, Volume 59, Issue 2, May 1998, pages 175 – 182.

Cochrane, J. (1988), How big is the random walk in GDP?, Journal of Political Economy, Volume 96, Issue 5, October 1988, pages 893 – 920.

Corbae, D and S. Ouliaris (1988), Cointegration and Tests of Purchasing Power Parity, The Review of Economics and Statistics, Volume 70, Issue 3, August 1988, pages 508 – 511.

Das, D. K (2001). Some Aspects of productivity growth and trade in Indian industry, Unpublished PhD Dissertation, Delhi School of Economics, University of Delhi.

Das, D. K. (2003), Quantifying Trade Barriers: Has Protection Declined Substantially in Indian Manufacturing, Economic and Political Weekly, January 31st Issue.

Dickey, D.A. and W.A. Fuller (1979), Distribution of the Estimators for Autoregressive Time Series With a Unit Root, Journal of the American Statistical Association, Volume 74, Issue 366, June 1979, pages 427 – 431.

Dickey, D.A. and W.A. Fuller (1981), Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root, Econometrica, Econometric Society, Volume 49, Issue 4, June 1981, pages 1057 – 1072.

Glynn, J., N. Perera and R. Verma (2007), Unit Root Tests and Structural Breaks: A Survey with Applications, Revista De Metodos Cuantitativos Para La Economia Y La Empressa, Volume 3, June 2007, pages 63 – 79.

Handbook of Indian Statistics (2001), Reserve Bank of India, Mumbai.

Lee, J. (undated), Gauss Codes for unit root tests, Culverhouse College of Commerce, The University of Alabama, http://cba.ua.edu/~jlee/gauss, accessed on June-July 2010.

Lee, J. and M. C. Strazicich (2001), Testing the Null of Stationarity in the Presence of a Structural Break, Applied Economics Letters, Taylor and Francis Journals, Volume 8, Issue 6, June 2001, pages 377 – 382.

Lee. J and M. C. Strazicich (2003), Minimum Lagrange Multiplier Unit Root Test with Two Structural Breaks, The Review of Economics and Statistics, MIT Press, Volume 85, Issue 4, November 2003, pages 1082 – 1089.

Lumsdaine, R. L. and D. H. Papell (1997), Multiple Trend Breaks and the Unit-Root Hypothesis, The Review of Economics and Statistics, MIT Press, Volume 79, Issue 2, May 1997, pages 212 – 218.

Maddala , G. S. and I. Kim (2003), Unit Roots, Cointegration, and Structural Change, Cambridge University Press, Fifth Edition, UK.

National Industrial Classification 1998, Central Statistical Organisation, Ministry of Statistics and Programme Implementation, Govt. of India, Nov 1998.

Nelson, C. R. and C. I. Plosser (1982), Trends and random walks in macroeconomic time series, Journal of Monetary Economics, Volume 10, Issue 2, 1982, pages 139 – 162.

Ng, S. and P. Perron (1997), Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power, Boston College Working Papers in Economics 369, Boston College Department of Economics, revised September 2000.

Ng, S. and P. Perron (2001), Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power, Econometrica, Econometric Society, Volume 69, Issue 6, November 2001, pages 1519 – 1554.

Nunes, L. C., P. Newbold and C. Kuan (1997), Testing for Unit Roots with Breaks: Evidence on the Great Crash and the Unit Root Hypothesis Reconsidered, Oxford Bulletin of Economics and Statistics, Department of Economics, University of Oxford, Volume 59, Issue 4, November 1997, pages 435 – 448.

OECD (2005), Data and Metadata Reporting and Presentation Handbook, OECD, Paris.

Ohara, H. I. (1999), A Unit Root test with multiple trend breaks: A Theory and An Application to US and Japanese Macroeconomic Time Series, The Japanese Economic Review, Volume 50, Issue 3, September 1999, pages 266 – 290.

Panagariya, A. (2004), India in the 1980s and 1990s: A Triumph of Reforms, IMF Working Paper No. WP/04/03, March.

Panagariya, A. (2007), Why India lags behind China and how it can bridge the gap, The World Economy, Volume 30, Issue 2, February 2007, pages 229 – 248.

Perron, P. (1988), Trends and random walks in macroeconomic time series : Further evidence from a new approach, Journal of Economic Dynamics and Control, Elsevier, Volume 12, Issue 2-3, June-September 1988, pages 297 – 332.

Perron, P. (1989), The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis, Econometrica, Econometric Society, Volume 57, Issue 6, November 1989, pages 1361 – 1401.

Perron, P. (1997), Further evidence on breaking trend functions in macroeconomic variables, Journal of Econometrics, Elsevier, Volume 80, Issue 2, October 1997, pages 355 – 385.

Perron, P. (2005), Dealing with Structural Breaks, Boston University - Department of Economics - Working Papers Series WP2005-017, Boston University, Department of Economics, April 2005.

Perron, P. and T. J. Vogelsang (1992), Nonstationarity and Level Shifts with an Application to Purchasing Power Parity, Journal of Business & Economic Statistics, American Statistical Association, Volume 10, Issue 3, July 1992, pages 301 – 320.

Phillips, P. C. B. (1986), Understanding spurious regressions in econometrics, Journal of Econometrics, Volume 33, Issue 3, December 1986, pages 311-340.

Phillips, P. C. B. and P. Perron (1988), Testing for a unit root in time series regression, Biometrika, Volume 75, Issue 2, June 1988, pages 335 – 346.

Schmidt, P. and P. C. B. Phillips (1992), LM Tests for a Unit Root in the Presence of Deterministic Trends, Oxford Bulletin of Economics and Statistics, Department of Economics, University of Oxford, Volume 54, Issue 3, August 1992, pages 257 – 287.

Sen, K. (2009), Trade policy, inequality and performance in Indian manufacturing, Routledge, Taylor & Francis Group, 2009.

Stulz, R. M. and W. Wasserfallen (1985), Macroeconomic time-series, business cycles and macroeconomic policies, Carnegie-Rochester Conference Series on Public Policy, Elsevier, Volume 22, Issue 1, January 1985, pages 9 – 53

Vogelsang, T. J. and P. Perron (1998), Additional Tests for a Unit Root Allowing for a Break in the Trend Function at an Unknown Time, International Economic Review, Department of Economics, University of Pennsylvania and Osaka University Institute of Social and Economic Research Association, Volume 39, Issue 4, November 1998, pages 1073 – 1100.

Wasserfallen, W. (1986), Non-Stationarities in Macro-Economic Time Series - Further Evidence and Implications, The Canadian Journal of Economics, Volume 19, Issue 3, August 1986, pages 498 – 510.

World Bank Statistical Manual, World Bank, permanent URL (available at the World Bank site) at: http://go.worldbank.org/OVSY1STWC0

Zivot, E. and D. W. K. Andrews (1992), Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis, Journal of Business & Economic Statistics, American Statistical Association, Volume 10, Issue 3, , July 1992, pages 251 – 270.

**Appendix Table1: National Industrial Classification 1998 at three digit level of industry aggregation, India**

| NIC-98 Code | Industry Description |
|---|---|
| 151 | Production, processing and preserving of meat, fish, fruits, veg., oils and fats |
| 152 | Manufacture of dairy product |
| 153 | Manufacture of grain mills products, starches and starch products and prepared animal feeds |
| 154 | Manufacture of other food products |
| 155 | Manufacture of beverages |
| 160 | Manufacture of tobacco products |
| 171 | Spinning, weaving and finishing of textiles |
| 172 | Manufacture of other textiles |
| 173 | Manufacture of knitted and crocheted fabrics and articles |
| 181 | Manufacturing of wearing apparel, except for fur apparel |
| 182 | Dressing and dyeing of fur, manufacture of articles of fur |
| 191 | Tanning and dressing of leather, manufacture of luggage hand bags, saddlery & harness |
| 192 | Manufacture of footwear |
| 201 | Saw milling and planing of wood |
| 202 | Manufacture of products of wood, cork, straw and plaiting materials |
| 210 | Manufacture of paper and paper product |
| 221 | Publishing |
| 222 | Printing and service activities related to printing |
| 231 | Manufacture of coke oven products |
| 232 | Manufactured refined petroleum products |
| 241 | Manufacture of basic chemicals |
| 242 | Manufacture of other chemical products |
| 243 | Manufacture of man-made fibres |
| 251 | Manufacture of rubber products |
| 252 | Manufacture of plastic products |
| 261 | Manufacture of glass and glass products |
| 269 | Manufacture of non-metallic mineral products n.e.c |
| 271 | Manufacture of basic iron and steel |
| 272 | Manufacture of basic precious and non-ferrous metals |
| 273 | Casting of metals |
| 281 | Manufacture of structural metal products, tanks, reservoirs and steam generators |

| | |
|---|---|
| **289** | Manufacture of other fabricated metal products, metal working service activities |
| **291** | Manufacture of general purpose machinery |
| **292** | Manufacture of special purpose machinery |
| **293** | Manufacture of domestic appliances, n.e.c |
| **300** | Manufacture of office, accounting and computer machinery |
| **311** | Manufacture of electric motors, generators and transformers |
| **312** | Manufacture of electricity distribution and control apparatus |
| **313** | Manufacture of insulated wire and cable |
| **314** | Manufacture of accumulators, primary cells and primary batteries |
| **315** | Manufacture of electric lamps and lighting equipment |
| **319** | Manufacture of other electrical equipment n.e.c |
| **321** | Manufacture of electronic valves and tubes and other electronic components |
| **322** | Manufacture of TV and radio transmitters and apparatus for line telephony and line telegraphy |
| **323** | Manufacture of TV and radio receivers, sound or video recording or reproducing apparatus, and associated goods |
| **331** | Manufacture of medical appliances and instruments and appliances for measuring, checking, testing, navigating and other purposes except optical instruments |
| **332** | Manufacture of optical instruments and photographic equipment |
| **333** | Manufacture of watches and clocks |
| **341** | Manufacture of motor vehicles |
| **342** | Manufacture of bodies for motor vehicles, trailer and semi trailers |
| **351** | Building and repair of ships and boats |
| **352** | Manufacture of railway and tramway locomotives and rollick stock |
| **353** | Manufacture of air craft and space craft |
| **359** | Manufacture of transport equipment n.e.c. |
| **361** | Manufacture of furniture |
| **369** | Manufacturing n.e.c |

**Source**:National Industrial Classification, 1998, CSO, MOSPI, Govt. of India